

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**

Gaila Bunsul  
7

Access DB#

# SEARCH REQUEST FORM

Scientific and Technical Information Center

59224

Requester's Full Name: 09/16/02, LEO Examiner #: 72967 Date: 1/29/02  
Art Unit: 1142 Phone Number 303 3455 Serial Number: 09/16/02 LEO  
Mail Box and Bldg/Room Location: CM/8AD3 Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need. MEJ

\*\*\*\*\*  
Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: Molecular Complexes which modify the immune system

Inventors (please provide full names): Schmuck, Tomaltian  
O'Hara

Earliest Priority Filing Date: 3/28/06

\*For Sequence Searches Only\* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

Please search

a composition comprising a cell wherein a  
molecular complex is bound to the surface of the cell,  
and the complex comprises  
4 fusion proteins.

two fusion proteins are <sup>Ig</sup> Heavy chain & a transmembrane  
polypeptide

and  
two fusion proteins are <sup>Ig</sup> light chain & a transmembrane  
polypeptide

21 Peptide linker See, Thaler GB

See, I 510,11 (Gly-Gly-Leu-Gly-Gly Ser)

and (Gly-Gly-Gly-Thr-Ser Gly

## STAFF USE ONLY

Searcher: Point of Contact:

Alex Wadlaw

Searcher Phone #: Technical Info. Specialist

Searcher Location: GM1-12C14 Tel: 308-4491

Date Searcher Picked Up: 7-11-02

Date Completed: 7-4-02

Searcher Prep & Review Time: \_\_\_\_\_

Clerical Prep Time: \_\_\_\_\_

Online Time: \_\_\_\_\_

## Type of Search

NA Sequence (#) \_\_\_\_\_

AA Sequence (#) 21

Structure (#) \_\_\_\_\_

Bibliographic ✓

Litigation \_\_\_\_\_

Fulltext \_\_\_\_\_

Patent Family \_\_\_\_\_

Other \_\_\_\_\_

## Vendors and cost where applicable

STN ✓

Dialog \_\_\_\_\_

Questel/Orbit \_\_\_\_\_

Dr.Link \_\_\_\_\_

Lexis/Nexis \_\_\_\_\_

Sequence Systems \_\_\_\_\_

WWW/Internet \_\_\_\_\_

Other (specify) \_\_\_\_\_

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 4, 2002, 08:00:16 : Search time 38.23 Seconds

(without alignments)  
11.625 Million cell updates/sec

Title: US-09-642-660-10

Perfect score: 33

Sequence: 1 GCGTSG 6

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 100 summaries

Database :

1: /SIDSR/gcgdata/geneseq/AA1980.DAT.\*  
2: /SIDSR/gcgdata/geneseq/AA1981.DAT.\*  
3: /SIDSR/gcgdata/geneseq/AA1982.DAT.\*  
4: /SIDSR/gcgdata/geneseq/AA1983.DAT.\*  
5: /SIDSR/gcgdata/geneseq/AA1984.DAT.\*  
6: /SIDSR/gcgdata/geneseq/AA1985.DAT.\*  
7: /SIDSR/gcgdata/geneseq/AA1986.DAT.\*  
8: /SIDSR/gcgdata/geneseq/AA1987.DAT.\*  
9: /SIDSR/gcgdata/geneseq/AA1988.DAT.\*  
10: /SIDSR/gcgdata/geneseq/AA1989.DAT.\*  
11: /SIDSR/gcgdata/geneseq/AA1990.DAT.\*  
12: /SIDSR/gcgdata/geneseq/AA1991.DAT.\*  
13: /SIDSR/gcgdata/geneseq/AA1992.DAT.\*  
14: /SIDSR/gcgdata/geneseq/AA1993.DAT.\*  
15: /SIDSR/gcgdata/geneseq/AA1994.DAT.\*  
16: /SIDSR/gcgdata/geneseq/AA1995.DAT.\*  
17: /SIDSR/gcgdata/geneseq/AA1996.DAT.\*  
18: /SIDSR/gcgdata/geneseq/AA1997.DAT.\*  
19: /SIDSR/gcgdata/geneseq/AA1998.DAT.\*  
20: /SIDSR/gcgdata/geneseq/AA1999.DAT.\*  
21: /SIDSR/gcgdata/geneseq/AA2000.DAT.\*  
22: /SIDSR/gcgdata/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	100.0	12	20	AAW95467
2	33	100.0	18	20	AAW95464
3	33	100.0	19	20	AAW95471
4	33	100.0	47	22	AAW3515
5	33	100.0	93	22	AAW17869
6	33	100.0	93	22	AAW30736
7	33	100.0	93	22	AAW05518
8	33	100.0	256	20	AAW06408
9	33	100.0	256	20	AAW95463
10	33	100.0	256	21	AAW9164
			22		AAW39721

12	33	100.0	256	22	AAW97366
13	33	100.0	264	22	AAW41507
14	33	100.0	272	22	AAW41789
15	33	100.0	272	22	AAW41790
16	33	100.0	300	21	AAW11577
17	33	100.0	302	21	AAW11576
18	33	100.0	303	21	AAW11575
19	33	100.0	348	20	AAW32858
20	33	100.0	348	22	AAW94509
21	33	100.0	378	19	AAW69428
22	33	100.0	499	21	AAW12987
23	33	100.0	499	21	AAW48191
24	33	100.0	520	22	AAW93590
25	33	100.0	552	21	AAW12986
26	33	100.0	552	21	AAW48190
27	33	100.0	557	22	AAW94160
28	33	100.0	582	22	AAW81993
29	33	100.0	602	22	AAW40004
30	33	100.0	625	22	AAW40003
31	33	100.0	678	22	AAW94603
32	33	100.0	2071	21	AAW84686
33	33	100.0	3025	22	AAW86196
34	30	90.9	32	20	AAW90042
35	30	90.9	66	19	AAW77518
36	30	90.9	81	21	AAW15636
37	30	90.9	83	21	AAW15635
38	30	90.9	95	20	AAW90054
39	30	90.9	116	21	AAW27159
40	30	90.9	256	21	AAW06217
41	30	90.9	258	22	AAW94584
42	30	90.9	291	21	AAW21313
43	30	90.9	291	21	AAW51418
44	30	90.9	315	21	AAW06216
45	30	90.9	346	21	AAW06215
46	30	90.9	358	22	AAW92530
47	30	90.9	372	22	AAW92109
48	30	90.9	410	22	AAW81172
49	30	90.9	430	21	AAW21312
50	30	90.9	430	21	AAW54147
51	30	90.9	453	21	AAW21311
52	30	90.9	453	21	AAW54146
53	30	90.9	473	22	AAW81727
54	30	90.9	482	22	AAW84911
55	30	90.9	537	21	AAW18564
56	30	90.9	538	18	AAW31300
57	30	90.9	538	19	AAW69393
58	30	90.9	538	19	AAW69397
59	30	90.9	538	21	AAW18552
60	30	90.9	538	21	AAW18560
61	30	90.9	543	22	AAW82506
62	30	90.9	543	21	AAW94201
63	30	90.9	543	22	AAW37672
64	30	90.9	560	17	AAW87529
65	30	90.9	594	21	AAW52451
66	30	90.9	594	21	AAW52479
67	30	90.9	679	22	AAW92241
68	30	90.9	1301	13	AAW28304
69	30	90.9	6	22	AAW75500
70	29	87.9	6	22	AAW67211
71	29	87.9	6	22	AAW76479
72	29	87.9	10	22	AAW94548
73	29	87.9	10	22	AAW94550
74	29	87.9	10	22	AAW97788
75	29	87.9	12	17	AAW87083
76	29	87.9	12	17	AAW92854
77	29	87.9	12	17	AAW92855
78	29	87.9	13	17	AAW92865
79	29	87.9	15	19	AAW51403
80	29	87.9	15	22	AAW36501
81	29	87.9	24	15	AAW54924
82	29	87.9	24	20	AAW05523
83	29	87.9	27	20	AAW13419
84	29	87.9	28	22	AAW84099

Human LUNX protein  
Human polypeptide  
Human polypeptide  
Human polypeptide  
Human polypeptide  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Human heat shock p  
Human protein sequ  
Human secreted pro  
Arabidopsis thalia  
Arabidopsis thalia  
Human protein sequ  
S. epidermidis ope  
Human polypeptide  
Human polypeptide  
Human protein sequ  
Amino acid sequenc  
HIV-1 subtype C pr  
B. steatothermophi  
Staphylococcus aur  
Arabidopsis thalia  
Arabidopsis thalia  
B. steatothermophi  
Zea mays protein f  
Arabidopsis thalia  
Human protein sequ  
Human protein sequ  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
C. glutamicum prote  
Mycobacterium tube  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
S. epidermidis ope  
Shrimp white spot  
Amino acid sequenc  
Pseudomonas fluore  
Pyrodictic gene  
Pyrodictic gene  
Amino acid sequenc  
Amino acid sequenc  
Pseudomonas fluore  
Fusarium venenatum  
Choline oxidase.  
Pseudomonas fluore  
Arabidopsis thalia  
Arabidopsis thalia  
C. glutamicum prote  
Acid alpha-amylase  
Peptide linker. S  
Peptide linker. S  
Linker used in Her  
Human complementar  
Human complementar  
Saccharomyces cere  
Cell adhesion modu  
Cell adhesion modu  
Stem cell factor r  
B43 scfv linker pe  
Hu gamma1 hinge +  
Flavivirus dimeric  
Conjugated peptide  
Immunomodulatory p

85	29	87.9	38	22	AAB84107	Immunomodulatory p
86	29	87.9	42	14	AAR3717	Plant non-specific
87	29	87.9	45	20	AAV19494	Amino acid sequenc
88	29	87.9	45	22	AAB67214	HER2 peptide ligand
89	29	87.9	45	22	AAB67218	TF peptide #3. Sy
90	29	87.9	53	19	AAV20790	Human neurofilamen
91	29	87.9	55	20	AAV05524	80kE zipper pepti
92	29	87.9	59	20	AAV05526	80kE zipper pepti
93	29	87.9	60	20	AAV05525	80kE zipper pept
94	29	87.9	89	21	AAB3137	Eucalyptus grandis
95	29	87.9	91	14	AAV18023	Hirudin fusion pro
96	29	87.9	93	22	AAV19220	Peptide #5654 enco
97	29	87.9	93	22	AAV1879	Peptide #5916 enco
98	29	87.9	98	17	AAR97234	MN protein region
99	29	87.9	100	21	AAV57712	Arabidopsis thalia
100	29	87.9	105	20	AAV13035	Human secreted pro

## ALIGNMENTS

## RESULT 1

ID AAW95467 standard; peptide: 12 AA.

AC AAW95467;

DT 29-MAR-1999 (first entry)

DE LS170 polypeptide fragment.

KW LS170 gene; In vivo imaging; Lung disease; cancer; drug screening.

OS Homo sapiens.

XX WO9856951-A1.

XX 17-DEC-1998.

XX 11-JUN-1998; 98WO-US11601.

XX 11-JUN-1997; 97US-0049183.

XX (ABBO ) ABBOTT LAB.

XX Billing-medel PA, Cohen M, Colpitts TL, Friedman PN;

XX Gordon J, Granadosen, Hodges SC, Klass MR, Kratochvil JD;

XX Roberts-rapp L, Russell JC, Stroupe SD;

XX WPI; 1999-060335/05.

XX New LS170 nucleic acid from lung tissue - useful for detecting,

XX monitoring, preventing and treating lung disease, particularly

XX cancer

XX Claim 17; Page 98; 120pp; English.

Sequences AAW95463-471 represent LS170 polypeptide fragments which contain at least one epitope in their sequences. The present invention relates to detection of a target LS170 polynucleotide that comprises a test sample with at least one LS170-specific nucleic acid that has at least 50 percent identity with any of the sequences (AAV0801-09) which are fragments derived from various clones of LS170 gene. The LS170 nucleic acid fragments represent a set of contiguous, partially overlapping sequences transcribed from lung tissue. They are used for diagnosis, staging, monitoring, in vivo imaging, prevention and treatment of lung disease, specifically cancer, and to indicate predisposition to such disease. Particularly detection of LS170 polynucleotide, LS170 antigens, or anti-LS170 antibodies is indicative of disease. Cells transformed with an expression system comprising the LS170 nucleic acid sequences are used to express recombinant polypeptides. The polypeptides are used to raise Ab and for drug screening. LS170-related nucleic acid can be used to isolate related sequences, as standards and reagents in

assays; as targets for drug screening, and as components or targets for therapy, e.g. as antisense, ribozyme or triplex-forming agents. Ab can be used to deliver therapeutic agents to LS170-expressing cells; directly as therapeutic agents (by neutralizing LS170 polypeptides); in competitive binding drug screens, and to generate anti-idiotypic antibodies for use in rational drug design.

SQ Sequence 12 AA;

Query Match 100.0%; Score 33; DB 20; Length 12;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGGTSG 6  
 Db 6 99GCSG 11

## RESULT 2

ID AAW95464 standard; peptide: 18 AA.

AC AAW95464;

DT 29-MAR-1999 (first entry)

DE LS170 polypeptide fragment.

KW LS170 gene; In vivo imaging; Lung disease; cancer; drug screening.

OS Homo sapiens.

XX WO9856951-A1.

XX 17-DEC-1998.

XX 11-JUN-1998; 98WO-US11601.

XX 11-JUN-1997; 97US-0049183.

XX (ABBO ) ABBOTT LAB.

XX Billing-medel PA, Cohen M, Colpitts TL, Friedman PN;

XX Gordon J, Granadosen, Hodges SC, Klass MR, Kratochvil JD;

XX Roberts-rapp L, Russell JC, Stroupe SD;

XX WPI; 1999-060335/05.

XX New LS170 nucleic acid from lung tissue - useful for detecting,

XX monitoring, preventing and treating lung disease, particularly

XX cancer

XX Claim 17; Page 97; 120pp; English.

Sequences AAW95463-471 represent LS170 polypeptide fragments which contain at least one epitope in their sequences. The present invention relates to detection of a target LS170 polynucleotide that comprises a test sample with at least one LS170-specific nucleic acid that has at least 50 percent identity with any of the sequences (AAV0801-09) which are fragments derived from various clones of LS170 gene. The LS170 nucleic acid fragments represent a set of contiguous, partially overlapping sequences transcribed from lung tissue. They are used for diagnosis, staging, monitoring, in vivo imaging, prevention and treatment of lung disease, specifically cancer, and to indicate predisposition to such disease. Particularly detection of LS170 polynucleotide, LS170 antigens, or anti-LS170 antibodies is indicative of disease. Cells transformed with an expression system comprising the LS170 nucleic acid sequences are used to express recombinant polypeptides. The polypeptides are used to raise Ab and for drug screening. LS170-related nucleic acid can be used to isolate related sequences, as standards and reagents in assays; as targets for drug screening, and as components or targets for therapy, e.g. as antisense, ribozyme or triplex-forming agents. Ab can be

CC used to deliver therapeutic agents to LS170-expressing cells; directly  
CC as therapeutic agents (by neutralising LS170 polypeptides); in  
CC competitive binding drug screens, and to generate anti-idiotypic  
CC antibodies for use in rational drug design.  
XX  
SQ Sequence 18 AA;

Query Match 100.0%; Score 33; DB 20; Length 18;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGGTSG 6  
| | | | |  
DB 7 gggtsq 12

## RESULT 3

AAM95471  
ID AAM95471 standard: peptide; 19 AA.

XX AAM95471;

XX 29-MAR-1999 (first entry)

DT LS170 polypeptide fragment.

XX LS170 gene; in vivo imaging; lung disease; cancer; drug screening.

XX Homo sapiens.

XX WO9556951-A1.

XX 17-DEC-1998.

PD 11-JUN-1998; 98WO-US11601.

XX 11-JUN-1997; 97US-0049183.

XX (ABBO ) ABBOTT LAB.

XX Billing-medel PA, Cohen M, Colpitts TL, Friedman PN;

PI Gordon J, Granadosen, Hodges SC, Klass MR, Kratochvil JD;

PI Roberts-rapp L, Russell JC, Stroupe SD;

XX WPI; 1999-060335/05.

XX New LS170 nucleic acid from lung tissue - useful for detecting,

XX monitoring, preventing and treating lung disease, particularly

XX cancer

XX Claim 17; Page 98; 120pp; English.

XX Sequences AAM95463-471 represent LS170 polypeptide fragments which  
CC contain at least one epitope in their sequences. The present invention  
CC relates to detection of a target LS170 polynucleotide that comprises  
CC treating a test sample with at least one LS170-specific nucleic acid that  
CC has at least 50 percent identity with any of the sequences (AAX0801-09)  
CC which are fragments derived from various clones of LS170 gene. The LS170  
CC nucleic acid fragments represent a set of contiguous, partially  
CC overlapping sequences transcribed from lung tissue. They are used for  
CC diagnosis, staging, monitoring, in vivo imaging, prevention and treatment  
CC of lung disease, specifically cancer, and to indicate predisposition to  
CC such disease. Particularly detection of LS170 polynucleotide, LS170  
CC antigens, or anti-LS170 antibodies is indicative of disease. Cells  
CC transformed with an expression system comprising the LS170 nucleic acid  
CC sequences are used to express recombinant polypeptides. The polypeptides  
CC are used to raise Ab and for drug screening. LS170-related nucleic acid  
CC sequences are used to isolate related sequences; as standards and reagents in  
CC assays; as targets for drug screening, and as components or targets for  
CC therapy, e.g. as antisense, ribozyme or triplex-forming agents. Ab can be  
CC used to deliver therapeutic agents to LS170-expressing cells; directly  
CC as therapeutic agents (by neutralising LS170 polypeptides); in

CC competitive binding drug screens, and to generate anti-idiotypic  
CC antibodies for use in rational drug design.  
XX  
SQ Sequence 19 AA;

Query Match 100.0%; Score 33; DB 20; Length 19;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGGTSG 6  
| | | | |  
DB 7 gggtsq 12

## RESULT 4

AAM33515  
ID AAM33515 standard: Protein; 47 AA.

XX AAM33515;

XX 17-OCT-2001 (first entry)

DE Peptide #7552 encoded by probe for measuring placental gene expression.

XX Probe; microarray; human; placenta; antenatal diagnosis;

XX genetic disorder.

XX Homo sapiens.

XX WO200157272-A2.

XX 09-AUG-2001.

PD 30-JAN-2001; 2001WO-US00663.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488897/53.

XX human genome-derived, single exon nucleic acid probes useful for

XX analyzing gene expression in human placenta -

XX Claim 27; SEQ ID No 33784; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENP;

XX see A113135-A157546). The present sequence is a peptide encoded by one

XX such probe. The probes are useful for producing a microarray for

XX predicting, measuring and displaying gene expression in samples derived

XX from human placenta. The probes are useful for antenatal diagnosis of

XX human genetic disorders.

XX Sequence 47 AA;

Query Match 100.0%; Score 33; DB 22; Length 47;  
Best Local Similarity 100.0%; Pred. No. 86;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGGTSG 6  
| | | | |  
DB 26 gggtsq 31

```

RESULT 5
AA17869
ID AAM17869 standard; Protein: 93 AA.
XX
AC AAM17869;
XX
DT 12-OCT-2001 (first entry)
XX
DE Peptide #4303 encoded by probe for measuring cervical gene expression.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
XX
KW cervical cancer.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PE 30-JAN-2001; 2001WO-US00670.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI: 2001-488901/53.
XX
PT human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells.
XX
PS Claim 27; SEQ ID NO 22695; 487bp; English.
XX
SQ
CC The present invention relates to human single exon nucleic acid probes
CC (SENP; see AAI10068-AA128459). The present sequence is a peptide encoded
CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC can be used to produce a single exon microarray, which can be used for
CC measuring human gene expression in a sample derived from human cervical
CC epithelial cells. By measuring gene expression, the probes are therefore
CC useful in grading and/or staging of diseases of the cervix, notably
CC cervical cancer.
CC
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 93 AA;

Query Match 100.0%; Score 33; DB 22; Length 93;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGTSG 6
   |||||
DB 46 gggts 51

RESULT 6
AAM30376
ID AAM30376 standard; Protein: 93 AA.
XX
AC AAM30376;
XX
DT 17-OCT-2001 (first entry)
XX

```

```

DE Peptide #4413 encoded by probe for measuring placental gene expression.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
XX
KW genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PE 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI: 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta.
XX
PS Claim 27; SEQ ID NO 30645; 654bp; English.
XX
SQ
CC The present invention relates to single exon nucleic acid probes (SENP;
CC see AAI3315-AA157546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders.
XX
SQ Sequence 93 AA;

Query Match 100.0%; Score 33; DB 22; Length 93;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGTSG 6
   |||||
DB 46 gggts 51

RESULT 7
AAM05518
ID AAM05518 standard; Protein: 93 AA.
XX
AC AAM05518;
XX
DT 09-OCT-2001 (first entry)
XX
DE Peptide #4200 encoded by probe for measuring breast gene expression.
XX
KW Probe; human; breast disease; breast cancer; development disorder;
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS Homo sapiens.
XX
PN WO200157270-A2.
XX
PD 09-AUG-2001.
XX
PE 29-JAN-2001; 2001WO-US00661.
XX
PR 04-FEB-2000; 2000US-0180312.

```

PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-476286/51.  
XX  
XX Novel single exon nucleic acid probe used to measuring gene expression  
XX in a human breast -  
XX  
XX Claim 27; SEQ ID No 14258; 322pp; English.  
XX  
XX The present invention relates to novel single exon nucleic acid probes  
XX (see AA100010-AA110067). The present sequence is a peptide encoded by one  
XX such probe. The probes are useful for measuring human gene expression in  
XX a human breast sample, where the probe hybridises at high stringency to a  
XX nucleic acid expressed in the human breast. The probes are useful for  
XX predicting, diagnosing, grading, staging, monitoring and prognosing  
XX diseases of the human breast, particularly those diseases with polygenic  
XX aetiology. The diseases include: breast cancer, disorders of development,  
XX inflammatory diseases of the breast, fibrocystic changes, proliferative  
XX breast disease and non-carcinoma tumours.  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 93 AA;  
  
Query Match 100.0%; Score 33; DB 22; Length 93;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 GCGTSG 6  
DB 46 gggtsq 51  
  
RESULT 8  
AA106408  
ID AAY06408 standard; Protein; 256 AA.  
XX  
XX AAY06408;  
XX  
XX 20-SEP-1999 (first entry)  
XX  
XX Human secreted protein nh796\_1.  
XX  
XX Secreted protein; nh796\_1; human; brain; thalamus; therapy;  
XX diagnosis.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
XX Peptide 7..19  
XX Protein /note="predicted leader/signal sequence"  
XX /note="mature protein"  
XX  
XX WO9935252-A2.  
XX  
XX 15-JUL-1999.  
XX  
XX 31-DEC-1998; 98WO-US27903.  
XX  
XX 30-DEC-1998; 98US-0222653.  
XX  
XX 02-JAN-1998; 98US-0070346.  
XX

XX  
XX (GENY) GENETICS INST INC.  
XX  
XX Agostino MJ, Collins-Racie LA, Jacobs K, LaVallie ER;  
XX McCoy JM, Werberg D, Steininger RJ, Treacy M;  
XX  
XX WPI; 1999-419350/35.  
XX  
XX N-PSDB; AAX59356.  
XX  
XX New polynucleotides encoding secreted human proteins  
XX  
XX Claim 25; Page 96-97; 100pp; English.  
XX  
XX This sequence represents a novel human secreted protein, termed  
XX nh796\_1. The sequence is predicted from a full-length cDNA clone  
XX (see AAX59356) isolated from a human adult brain (thalamus) cDNA  
XX library. The invention provides cDNA clones (see AAX59352-58)  
XX encoding novel secreted proteins (see AAY06404-10) of the human  
XX testis, brain and foetal kidney. The polynucleotides and  
XX proteins are predicted to have biological activities which would  
XX make them suitable for treating, preventing or ameliorating medical  
XX conditions in humans and animals, although no supporting data are  
XX provided. Suggested activities include nutritional, cytokine, cell  
XX proliferation or differentiation, immunostimulant (e.g. as  
XX vaccine) or immunosuppressive, haematopoiesis regulating, tissue  
XX growth, activin or inhibin, chemotactic or chemokinetic,  
XX haemostatic, thrombolytic, receptor/ligand, antiinflammatory,  
XX catheterin or tumour invasion suppressor, and tumour inhibition  
XX activity.  
XX  
SQ Sequence 256 AA;  
  
Query Match 100.0%; Score 33; DB 20; Length 256;  
Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 GCGTSG 6  
DB 80 gggtsq 85  
  
RESULT 9  
AAM95463  
ID AAM95463 standard; Protein; 256 AA.  
XX  
XX AAM95463;  
XX  
XX 29-MAR-1999 (first entry)  
XX  
XX LSI70 polypeptide sequence.  
XX  
XX LSI70 gene; in vivo imaging; lung disease; cancer; drug screening.  
XX  
XX Homo sapiens.  
XX  
XX WO9856951-A1.  
XX  
XX 17-DEC-1998.  
XX  
XX 11-JUN-1998; 98WO-US11601.  
XX  
XX 11-JUN-1997; 97US-0049183.  
XX  
XX (ABBO) ABBOTT LAB.  
XX  
XX Billing-medel PA, Cohen M, Colpitts TL, Friedman PN;  
XX Gordon J, Granadosen, Hodges SC, Klass MR, Kratochvil JD;  
XX Roberts-rapp L, Russell JC, Stroupe SD;  
XX  
XX WPI; 1999-060335/05.  
XX  
XX N-PSDB; AAX00809.  
XX

PT New Ls170 nucleic acid from lung tissue - useful for detecting,  
PT monitoring, preventing and treating lung disease, particularly  
PT cancer  
XX  
XX  
PS Claim 17: Page 96-97; 120pp; English.  
XX  
CC Sequences AAW95463-471 represent Ls170 polypeptide fragments which  
CC contain at least one epitope in their sequences. The present invention  
CC relates to detection of a target Ls170 polynucleotide that comprises  
CC treating a test sample with at least one Ls170-specific nucleic acid that  
CC has at least 50 percent identity with any of the sequences (AAW0801-09)  
CC which are fragments derived from various clones of Ls170 gene. The Ls170  
CC nucleic acid fragments represent a set of contiguous, partially  
CC overlapping sequences transcribed from lung tissue. They are used for  
CC diagnosis, staging, monitoring, in vivo imaging, prevention and treatment  
CC of lung disease, specifically cancer, and to indicate predisposition to  
CC such disease. Particularly detection of Ls170 polynucleotide, Ls170  
CC antigens, or anti-Ls170 antibodies is indicative of disease. Cells  
CC transformed with an expression system comprising the Ls170 nucleic acid  
CC sequences are used to express recombinant polypeptides. The polypeptides  
CC are used to raise Ab and for drug screening. Ls170-related nucleic acid  
CC can be used to isolate related sequences; as standards and reagents in  
CC assays; as targets for drug screening, and as components or targets for  
CC therapy, e.g. as antisense, ribozyme or triplex-forming agents. Ab can be  
CC used to deliver therapeutic agents to Ls170-expressing cells; directly  
CC as therapeutic agents (by neutralising Ls170 polypeptides); in  
CC competitive binding drug screens, and to generate anti-idiotypic  
CC antibodies for use in rational drug design.  
CC  
XX  
SQ Sequence 256 AA;  
  
Query Match 100.0%; Score 33; DB 20; Length 256;  
Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GCGTSG 6  
| | | | |  
Db 80 99gts 85  
  
RESULT 10  
AAV69164  
ID AAV69164 standard; Protein: 256 AA.  
XX  
AC AAV69164;  
XX  
DT 30-MAY-2000 (first entry)  
XX  
DE Amino acid sequence of lung specific protein Lng110.  
XX  
XX Lung specific gene; lung cancer; metastatic lung cancer; Imaging.  
XX  
OS Homo sapiens.  
XX  
PN W0200008206-A1.  
XX  
PD 17-FEB-2000.  
XX  
PF 19-JUL-1999; 99WO-US16247.  
XX  
PR 04-AUG-1998; 98US-0095233.  
XX  
PA (DIAD-) DIADEXUS LLC.  
XX  
PI Yang F, Sun Y, Recipon H, Macina RA.  
XX  
DR WPI; 2000-195589/17.  
XX  
DR N-PSDB: AAV61165.  
XX  
PT Lung specific genes for developing products for diagnosing, monitoring,  
PT staging, prognosticating, imaging and treating lung cancer -  
XX

PS Example 2: Page 35-36; 37pp; English.  
XX  
CC The present sequence is encoded by a lung specific gene, designated  
CC Lng110, clone ID 1355520, gene ID 236760. The specification describes  
CC methods for diagnosing the presence of lung cancer in a patient. The  
CC method comprises measuring levels of lung specific genes in cells.  
CC Tissues or bodily fluids, and comparing the level to that of a normal  
CC human control. The methods can be used for diagnosing, monitoring,  
CC staging or prognosticating lung cancer, especially metastatic lung  
CC cancer. Antibodies against proteins encoded by lung specific genes  
CC can be used for imaging or, when conjugated to a cytotoxic agent, for  
CC treating lung cancer.  
XX  
SQ Sequence 256 AA;  
  
Query Match 100.0%; Score 33; DB 21; Length 256;  
Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GCGTSG 6  
| | | | |  
Db 80 99gts 85  
  
RESULT 11  
AAM39721  
ID AAM39721 standard; Protein: 256 AA.  
XX  
AC AAM39721;  
XX  
DT 22-OCT-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 2866.  
XX  
KW Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia.  
XX  
XX Homo sapiens.  
XX  
OS  
XX  
PN W020015312-A1.  
XX  
PD 26-JUL-2001.  
XX  
PE 26-DEC-2000; 2000WO-US34263.  
XX  
PR 21-JAN-2000; 2000US-0488725.  
XX  
PR 25-APR-2000; 2000US-0552317.  
XX  
PR 09-JUL-2000; 2000US-0598042.  
XX  
PR 19-JUL-2000; 2000US-0620312.  
XX  
PR 03-AUG-2000; 2000US-0653450.  
XX  
PR 14-SEP-2000; 2000US-0662191.  
XX  
PR 19-OCT-2000; 2000US-0693036.  
XX  
PR 29-NOV-2000; 2000US-0727344.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
XX  
DR WPI; 2001-442253/47.  
XX  
DR N-PSDB: AAI58877.  
XX  
PT Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -  
XX  
XX Example 4; SEQ ID NO 2866; 10078pp; English.  
XX

CC The invention relates to human nucleic acids (AA157798-AA161369) and  
CC the encoded polypeptides (AA038642-AA042213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
SQ Sequence 256 AA;

Query Match 100.0%; Score 33; DB 22; Length 256;  
Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGTSG 6  
|||||  
DB 80 999ts9 85

## RESULT 12

AA097366  
ID AA097366 standard; protein; 256 AA.

XX  
AC AA097366;

XX  
DT 15-AUG-2001 (first entry)

XX  
DE Human LUNX protein.

XX  
KW LUNX; human; cancer; micrometastatic cancer.

XX  
OS Homo sapiens.

XX  
PN JP2001078772-A.

XX  
PD 27-MAR-2001.

XX  
PF 07-SEP-1999; 99JP-0253186.

XX  
PR 07-SEP-1999; 99JP-0253186.

XX  
PA (SAKA ) OTSUKA PHARM CO LTD.

XX  
WIPI: 2001-313367/33.

XX  
DR N-PSDB; AA027756, AA027757.

XX  
PT Polynucleotide encoding LUNX gene product useful for the detection of  
PT cancer especially micrometastatic cancer -

XX  
PS Claim 1; Page 25; 30pp; Japanese.

XX  
CC This invention relates to the human LUNX protein and the polynucleotide  
CC sequence encoding it. The invention includes a vector containing a LUNX  
CC polynucleotide, a host cell transformed with the vector, and an antibody  
CC that binds to LUNX. The gene can be used for cancer diagnosis and  
CC diagnosis of micrometastatic cancer and for the production of the LUNX  
CC gene product. The present sequence represents human LUNX.

XX  
SQ Sequence 256 AA;

Query Match 100.0%; Score 33; DB 22; Length 256;  
Best Local Similarity 100.0%; Pred. No. 3.8e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGTSG 6  
|||||  
DB 80 999ts9 85

## RESULT 13

AA041507  
ID AA041507 standard; Protein; 264 AA.

XX  
AC AA041507;

XX  
DT 22-OCT-2001 (first entry)

XX  
DE Human polypeptide SEQ ID NO 6438.

XX  
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;

XX  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;

XX  
KW leukaemia.

XX  
OS Homo sapiens.

XX  
PN WO200153312-A1.

XX  
PD 26-JUL-2001.

XX  
PF 26-DEC-2000; 2000WO-US34263.

XX  
PR 21-JAN-2000; 2000US-0488725.

XX  
PR 25-APR-2000; 2000US-0552317.

XX  
PR 09-JUL-2000; 2000US-0598042.

XX  
PR 19-JUL-2000; 2000US-0620312.

XX  
PR 03-AUG-2000; 2000US-0653450.

XX  
PR 14-SEP-2000; 2000US-0662191.

XX  
PR 19-OCT-2000; 2000US-0693036.

XX  
PR 29-NOV-2000; 2000US-0727344.

XX  
PA (HYSE-) HYSEQ INC.

XX  
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX  
WIPI: 2001-442253/47.

XX  
DR N-PSDB; AA160663.

XX  
PT Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -

XX  
PS Example 2; SEQ ID NO 6438; 10078pp; English.

XX  
CC The invention relates to human nucleic acids (AA157798-AA161369) and  
CC the encoded polypeptides (AA038642-AA042213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.

XX  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.

XX  
SQ Sequence 264 AA;

Query Match 100.0%; Score 33; DB 22; Length 264;  
Best Local Similarity 100.0%; Pred. No. 3,9e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
|||||  
Db 88 ggtsg 93

## RESULT 14

AA041789  
ID AA041789 standard; Protein: 272 AA.

AA041789;

22-OCT-2001 (first entry)

Human polypeptide SEQ ID NO 6720.

Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;  
peripheral nervous system; neuropathy; central nervous system; CNS;  
Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
chemokine; thrombolytic; drug screening; arthritis; inflammation;  
leukaemia.

Homo sapiens.

WO200153312-A1.

26-JUL-2001.

26-DEC-2000; 2000MO-US34263.

21-JAN-2000; 2000US-0488725.

25-APR-2000; 2000US-0552317.

09-JUL-2000; 2000US-0598042.

19-JUL-2000; 2000US-0620312.

03-AUG-2000; 2000US-0653450.

14-SEP-2000; 2000US-0662191.

19-OCT-2000; 2000US-0693036.

29-NOV-2000; 2000US-0727344.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
Zhao QA, Zhou P, Goodrich R, Drmanac RT;

WPI: 2001-442253/47.

N-PSDB; AA160945.

Novel nucleic acids and polypeptides, useful for treating disorders

such as central nervous system injuries -

Example 2: SEQ ID NO 6720; 10078pp; English.

The invention relates to human nucleic acids (AA157798-AA161369) and  
the encoded polypeptides (AA038642-AA042213) with noctropic,  
immunosuppressant and cytostatic activity. The polynucleotides are useful  
in gene therapy. A composition containing a polypeptide or polynucleotide  
system, such as peripheral nervous injuries, peripheral neuropathy and  
localised neuropathies and central nervous system diseases, such as  
Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
utilisation of the activities such as: immune system suppression,  
Activin/Inhibin activity, chemotactic/chemokinetic activity, haemostatic  
and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
assays for receptor activity, arthritis and inflammation, leukaemias and  
C.N.S disorders.

Note: The sequence data for this patent did not form part of the printed  
specification.

Sequence 272 AA;

Query Match 100.0%; Score 33; DB 22; Length 272;  
Best Local Similarity 100.0%; Pred. No. 4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
|||||  
Db 123 ggtsg 128

## RESULT 15

AA041790  
ID AA041790 standard; Protein: 272 AA.

AA041790;

22-OCT-2001 (first entry)

Human polypeptide SEQ ID NO 6721.

Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;  
peripheral nervous system; neuropathy; central nervous system; CNS;  
Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
chemokine; thrombolytic; drug screening; arthritis; inflammation;  
leukaemia.

Homo sapiens.

WO200153312-A1.

26-JUL-2001.

26-DEC-2000; 2000MO-US34263.

21-JAN-2000; 2000US-0488725.

25-APR-2000; 2000US-0552317.

09-JUL-2000; 2000US-0598042.

19-JUL-2000; 2000US-0620312.

03-AUG-2000; 2000US-0653450.

14-SEP-2000; 2000US-0662191.

19-OCT-2000; 2000US-0693036.

29-NOV-2000; 2000US-0727344.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
Zhao QA, Zhou P, Goodrich R, Drmanac RT;

WPI: 2001-442253/47.

N-PSDB; AA160946.

Novel nucleic acids and polypeptides, useful for treating disorders

such as central nervous system injuries -

Example 2: SEQ ID NO 6721; 10078pp; English.

The invention relates to human nucleic acids (AA157798-AA161369) and  
the encoded polypeptides (AA038642-AA042213) with noctropic,  
immunosuppressant and cytostatic activity. The polynucleotides are useful  
in gene therapy. A composition containing a polypeptide or polynucleotide  
system, such as peripheral nervous injuries, peripheral neuropathy and  
localised neuropathies and central nervous system diseases, such as  
Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
utilisation of the activities such as: immune system suppression,



CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.

XX  
SO Sequence 272 AA:

Query Match 100.0%; Score 33; DB 22; Length 272;  
Best Local Similarity 100.0%; Pred. No. 4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGCTSG 6  
Db 123 gggctsg 128

RESULT 16

AG1577  
ID AG1577 standard; Protein; 300 AA.

XX  
AC AG1577;

XX  
DT 17-OCT-2000 (first entry)

XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 10345.

XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX  
OS Arabidopsis thaliana.

XX  
PN EPI033405-A2.

XX  
PD 06-SEP-2000.

XX  
PF 25-FEB-2000; 2000EP-0301439.

XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123568.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139869.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.

PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 23-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 25-AUG-1999; 99US-0149930.  
PR 26-AUG-1999; 99US-0150566.  
PR 27-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 31-AUG-1999; 99US-0151303.  
PR 01-SEP-1999; 99US-0151338.  
PR 07-SEP-1999; 99US-0151930.  
PR 10-SEP-1999; 99US-0152363.  
PR 13-SEP-1999; 99US-0153070.  
PR 15-SEP-1999; 99US-0153758.  
PR 16-SEP-1999; 99US-0154018.  
PR 20-SEP-1999; 99US-0154039.  
PR 22-SEP-1999; 99US-0154779.  
PR 23-SEP-1999; 99US-0155139.  
PR 24-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156456.  
PR 29-SEP-1999; 99US-0156536.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157173.  
PR 06-OCT-1999; 99US-0157753.  
PR 07-OCT-1999; 99US-0157865.  
PR 08-OCT-1999; 99US-0158029.  
PR 12-OCT-1999; 99US-0158358.  
PR 13-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159283.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.

PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
  
Query Match 100.0%; Score 33; DB 21; Length 300;  
Best Local Similarity 100.0%; Pred. No. 4,4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GGTSG 6  
DB 11 ggtsg 16  
  
RESULT 17  
AAG1576  
ID AAG1576 standard; Protein: 302 AA.  
XX  
AC AAG1576;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 10344.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP103405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 28-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134376.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145276.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 28-JUL-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.

PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149375.  
PR 18-AUG-1999; 99US-0149426.  
PR 18-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 23-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154038.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155486.  
PR 28-SEP-1999; 99US-0156456.  
PR 29-SEP-1999; 99US-0156599.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0158293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 18-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 100.0%; Score 33; DB 21; Length 302;  
Best Local Similarity 100.0%; Pred. No. 4.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGCTSG 6  
Db 13 g99tsg 18  
RESULT 18  
AAG11575  
ID AAG11575 standard; Protein: 303 AA.  
XX AAG11575;  
XX 17-OCT-2000 (first entry)  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 10343.  
DE Protein identification; signal transduction pathway; metabolic pathway;  
KM hydrolisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX Arabidopsis thaliana.  
OS EP103405-A2.  
PN  
XX  
XX 06-SEP-2000.  
FD  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
PF  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 04-MAY-1999; 99US-0132408.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 19-MAY-1999; 99US-0134768.  
PR 20-MAY-1999; 99US-0134941.  
PR 21-MAY-1999; 99US-0135124.  
PR 24-MAY-1999; 99US-0135353.  
PR 25-MAY-1999; 99US-0135629.  
PR 27-MAY-1999; 99US-0136021.  
PR 28-MAY-1999; 99US-0136592.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138847.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139763.  
PR 22-JUN-1999; 99US-0139817.  
PR 23-JUN-1999; 99US-0139859.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140921.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 20-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 26-JUL-1999; 99US-0145224.  
PR 27-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 28-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 09-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148365.  
PR 13-AUG-1999; 99US-0148684.

```

PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Query Match 100.0%; Score 33; DB 21; Length 303;
Best Local Similarity 100.0%; Pred. No. 44e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGGTSG 6
DB 14 gggtsag 19

```

RESULT 19  
AAV32858

```

ID AAV32858 standard; Protein; 348 AA.
XX
AC AAV32858;
XX
DT 01-NOV-1999 (first entry)
XX
DE Human heat shock protein homologue 1 (HSPH-1) protein sequence.
XX
KW Human heat shock protein homologue; HSPH-1; HSPH-2; diagnosis; cancer;
inflammation; AIDS; Crohn's disease; allergy; therapy; drug screening.
XX
OS Homo sapiens.
XX
PN US5945287-A.
XX
PD 31-AUG-1999.
XX
PE 19-NOV-1997; 97US-0974546.
XX
PR 19-NOV-1997; 97US-0974546.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Corley NC, Hillman JL, Lal P, Shah P;
XX
DR WPI; 1999-526253/44.
XX
PT N-PSDB; AA211066.
XX
PS New molecular chaperones useful for diagnosis, prevention and
treatment of cancer and inflammation
XX
SQ Claim 1; Fig 1; 35pp; English.
XX
CC This sequence is the human heat shock protein homologue, HSPH-1 of
the invention. The HSPH-1 polynucleotide was isolated from the BRA1UT21
cDNA library constructed from cancerous brain tissue. The HSPH-2 DNA
sequence was isolated from the BRSNOT18 cDNA library constructed from
cancerous breast tissue. The HSPH polynucleotides and their complements
are useful for diagnosis of conditions or disorders associated with HSPH
expression. HSPH complements are also useful as antagonists for
prevention or treatment of cancer e.g. adenocarcinoma, breast and skin,
and inflammation associated with e.g. AIDS, Crohn's disease and
allergies. HSPH-1 and HSPH-2 are useful for drug screening using
CC libraries of compounds. The HSPH-1, HSPH-2 nucleotide sequences, and
CC their complements, are useful for detecting upstream sequences
CC e.g. promoters and regulatory elements. Vectors containing the DNA
sequences are useful for the preparation of HSPH proteins and mRNA
probes.
XX
SQ Sequence 348 AA;

```

```

Query Match 100.0%; Score 33; DB 20; Length 348;
Best Local Similarity 100.0%; Pred. No. 5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGGTSG 6
DB 75 gggtsag 80

```

```

RESULT 20
AAB94509
ID AAB94509 standard; Protein; 348 AA.
XX
AC AAB94509;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:15218.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX

```

OS Homo sapiens.  
 XX  
 PN EP1074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-0116126.  
 XX  
 PR 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 DR WPI: 2001-318749/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 PS Claim 8: SEQ ID 15218; 2537bp + CD ROM, English.  
 XX  
 CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 XX  
 SQ Sequence 348 AA:  
 XX  
 Query Match 100.0%; Score 33; DB 22; Length 348;  
 Best Local Similarity 100.0%; Pred. No. 5e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GCGTSG 6  
 XX |  
 XX |  
 XX |  
 XX |  
 XX |  
 DB 75 ggggtsg 80  
 XX  
 RESULT 21  
 AAM69428  
 ID AAM69428 standard; Protein: 378 AA.  
 XX  
 AC AAM69428;  
 XX  
 DT 18-JAN-1999 (first entry)  
 XX  
 DE Human secreted protein bp537\_4.  
 XX

KM Secreted protein; human; bp537\_4.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO9840404-A2.  
 XX  
 PD 17-SEP-1998.  
 XX  
 PF 09-MAR-1998; 98WO-US04601.  
 XX  
 PR 06-MAR-1998; 98US-0036321.  
 PR 11-MAR-1997; 97US-0815381.  
 XX  
 PA (GEMV ) GENETICS INST INC.  
 XX  
 PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;  
 PI Racine LA, Spaulding V, Treacy M;  
 XX  
 DR WPI: 1998-520802/44.  
 DR N-PSDB: AAV58758.  
 XX  
 PT New isolated polynucleotides and secreted proteins - useful as, e.g.  
 PT nutritional additives, immunostimulators, haemotopoiesis regulators  
 PT and as diagnostic agents  
 XX  
 PS Claim 24: Page 79-80; 119pp; English.  
 XX  
 CC This polypeptide, designated bp537\_4, comprises a novel human  
 CC secreted protein that is encoded by a cDNA clone (see AAV58758)  
 CC isolated from a human foetal kidney cDNA library. Homology is  
 CC shown to some database sequences. The invention provides isolated  
 CC polynucleotides (see AAV58754-63) obtained from human adult testis,  
 CC brain, retina or placenta, or from foetal kidney or brain cDNA  
 CC libraries. These are all deposited as ATCC 98353. They encode  
 CC novel human secreted proteins (see AAM69423-33) that may have e.g.  
 CC nutritional activity, immune stimulating or suppressing activity,  
 CC haematopoiesis regulating activity, tissue growth activity,  
 CC actinin/inhibin activity, chemotactic/chemokinetic activity,  
 CC haemostatic and thrombolytic activity, receptor/ligand activity,  
 CC antiinflammatory activity, cadherin/tumour invasion suppressor  
 CC activity, tumour inhibition or other activities. They may also be  
 CC used for diagnostic purposes.  
 XX  
 SQ Sequence 378 AA:  
 XX  
 Query Match 100.0%; Score 33; DB 19; Length 378;  
 Best Local Similarity 100.0%; Pred. No. 5.4e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GCGTSG 6  
 XX |  
 XX |  
 XX |  
 XX |  
 XX |  
 DB 252 ggggtsg 257  
 XX  
 RESULT 22  
 AAG12987  
 ID AAG12987 standard; Protein: 499 AA.  
 XX  
 AC AAG12987;  
 XX  
 DT 17-OCT-2000 (first entry)  
 XX  
 DE Arabidopsis thaliana protein fragment SRO ID NO: 12312.  
 XX  
 KM Protein identification: signal transduction pathway; metabolic pathway;  
 KM hybridisation assay; genetic mapping; gene expression control; promoter;  
 KM termination sequence.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 PN EP1033405-A2.  
 XX

PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123160.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139817.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139889.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147202.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0148684.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.

PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155639.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161932.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 100.0%; Score 33; DB 21; Length 499;  
Best Local Similarity 100.0%; Pred. No. 6,9e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGTSG 6  
|||||  
Db 25 ggcTcg 30

## RESULT 23

AAC48191

ID AAC48191 standard: Protein: 499 AA.

XX AAC48191:

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 60832.

KW Protein identification: signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136382.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137528.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.



PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 28-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147419.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151338.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.

PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158329.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 100.0%; Score 33; DB 21; Length 499;  
Best Local Similarity 100.0%; Pred. No. 6; 9e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCTSG 6  
|||||  
Db 25 gsgtsg 30

RESULT 24.  
AAB93590  
ID AAB93590 standard; Protein: 520 AA.  
XX AC AAB93590;  
XX 26-JUN-2001 (first entry)  
XX DT  
XX DE Human protein sequence SEQ ID NO:13016.  
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX OS Homo sapiens.  
XX PN EP1074617-A2.  
XX PD  
XX PD 07-FEB-2001.  
XX PF 28-JUL-2000; 2000EP-0116126.  
XX PR 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX  
PA (HELI-) HELIX RES INST.  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
XX full-length cDNAs -  
PS Claim 8; SEQ ID 13016; 2537p + CD ROM; English.  
CC The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
SQ Sequence 520 AA:  
  
Query Match 100.0%; Score 33; DB 22; Length 520;  
Best Local Similarity 100.0%; Pred. No. 7.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 GCGTSG 6  
|||||  
Db 501 ggcgtsg 506  
  
RESULT 25  
AAG12986  
ID AAG12986 standard; Protein; 552 AA.  
XX  
AC AAG12986;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 12311.  
XX  
XX  
KM Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.  
OS Arabidopsis thaliana.  
OS  
XX  
XX  
PN EP1033405-A2.  
PD  
PD 06JSEP-2000.  
XX  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139859.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142054.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144332.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 04-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149930.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157573.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158332.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.

PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 22-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 100.0%; Score 33; DB 21; Length 552;  
Best Local Similarity 100.0%; Pred. No. 7.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Caps 0;

Oy 1 GCGTSG 6  
|||||  
Db 78 gggtsq 83

RESULT 26  
AACG48190  
ID AACG48190 standard; Protein; 552 AA.  
XX AC  
XX AAG48190;  
XX DT 18-OCT-2000 (first entry)  
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 60831.  
XX KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX OS Arabidopsis thaliana.  
XX PN  
XX EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-0301439.  
XX XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 23-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0132856.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139467.  
PR 18-JUN-1999; 99US-0139468.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140352.  
PR 23-JUN-1999; 99US-0140358.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142927.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0146389.  
PR 04-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 10-AUG-1999; 99US-0147935.  
PR 11-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148555.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151920.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.

PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 21-OCT-1999; 99US-0160980.  
PR 21-OCT-1999; 99US-0160981.  
PR 21-OCT-1999; 99US-0160989.  
PR 21-OCT-1999; 99US-0161404.  
PR 21-OCT-1999; 99US-0161405.  
PR 21-OCT-1999; 99US-0161406.  
PR 21-OCT-1999; 99US-0161359.  
PR 21-OCT-1999; 99US-0161360.  
PR 21-OCT-1999; 99US-0161361.  
PR 21-OCT-1999; 99US-0161920.  
PR 21-OCT-1999; 99US-0161992.  
PR 21-OCT-1999; 99US-0161993.  
PR 21-OCT-1999; 99US-0162142.

Query Match 100.0%; Score 33; DB 21; Length 552;  
Best Local Similarity 100.0%; Pred. No. 7.5e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
|||||  
DB 78 99tsg 83

RESULT 27

AA894160  
ID AAB94160 standard; Protein; 557 AA.

AC AAB94160;

DT 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:14452.

DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.

OS Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 09-JUN-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

PA (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

DR WPI: 2001-318749/34.

PR Primer sets for synthesizing polynucleotides, particularly the 5602

PR full-length cDNAs defined in the specification, and for the detection

PR full-length cDNAs -

PR Claim 8; SEQ ID 14452; 2537pp + CD ROM; English.

CC The present invention describes primer sets for synthesizing 5602

CC full-length cDNAs defined in the specification, where a primer set

CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary

CC to the complementary strand of a polynucleotide which comprises one of

CC the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a

CC polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and

CC in gene therapy. The primers are useful for synthesizing polynucleotides,

CC particularly full-length cDNAs. The primers are also useful for the

CC detection and/or diagnosis of the abnormality of the proteins encoded by

CC the full-length cDNAs. The primers allow obtaining of the full-length

CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to

CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632

CC represent oligonucleotides, all of which are used in the exemplification

CC of the present invention.

SO Sequence 557 AA;

Query Match 100.0%; Score 33; DB 22; Length 557;  
Best Local Similarity 100.0%; Pred. No. 7.6e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
|||||  
DB 431 99tsg 436

RESULT 28

AA81993  
ID AAG81993 standard; Protein; 582 AA.

AC AAG81993;

DT 03-SEP-2001 (first entry)

DE S. epidermidis open reading frame protein sequence SEQ ID NO:1080.

DE Staphylococcus epidermidis SRI strain; infection; diagnosis;

DE vaccination; endocarditis.

OS Staphylococcus epidermidis.

PN WO200134809-A2.

PD 17-MAY-2001.

PF 09-NOV-2000; 2000WO-US30782.

PR 09-NOV-1999; 99US-0164258.

PR (GLAX ) GLAXO GROUP LTD.

PI Kimmerly WJ;

PI WPI: 2001-316495/33.

DR N-PSDB; AAH52843.

DR Nucleic acids encoding polypeptides from Staphylococcus epidermidis,

DR useful for vaccinating against infections, e.g. endocarditis -

DR Claim 18; Page 315; 2188pp; English.

CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides

CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.

CC (I) and (II) can have antibacterial activity and therefore can be used

CC in vaccination. The nucleic acids (I) may be used to produce the

CC S. epidermidis polypeptides (II) via the production of vectors

CC containing them which are used to produce hosts cells which express the

CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be

CC	used to vaccinate subjects and to raise antibodies against the bacteria.
CC	The polypeptides may also be used to assay for other inhibitors of their
CC	activity and therefore identify compounds that may be used for the
CC	treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC	AAH5090 represent specifically claimed S. epidermidis genomic DNA
CC	polynucleotide sequences from the present invention. AAH5091 to
CC	AAH5098 represent oligonucleotide sequences and primers which are used
CC	in the exemplification of the present invention.
CC	N.B. The present invention specifically claims all the polynucleotide
CC	sequences given in the sequence listing of the present specification,
CC	however the sequence listing only goes up to SEQ ID NO:4454 so even
CC	though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC	no sequences are present for SEQ ID NO:4455 to 4464.
CC	
XX	
SO	Sequence 582 AA:
QY	1 GGGTSG 6
Db	488 99tsg 493
Query Match	100.0%; Score 33; DB 22; Length 582;
Best Local Similarity	100.0%; Pred. No. 7.9e+02;
Matches 6; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
RESULT 29	
AAH40004	AAH40004 standard; Protein: 602 AA.
AAH40004;	
22-OCT-2001 (first entry)	
Human polypeptide SEQ ID NO 3149.	
Human: neutrotic; immunosuppressant; cytosatic; gene therapy; cancer;	
peripheral nervous system; neuropathy; central nervous system; CNS;	
Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;	
amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;	
chemokine; thrombolytic; drug screening; arthritis; inflammation;	
leukemia.	
Homo sapiens.	
WO200153312-A1.	
26-JUL-2001.	
26-DEC-2000; 2000WO-US34263.	
21-JAN-2000; 2000US-0488725.	
25-APR-2000; 2000US-0552317.	
09-JUL-2000; 2000US-0598042.	
19-JUL-2000; 2000US-0620312.	
03-AUG-2000; 2000US-0653450.	
14-SEP-2000; 2000US-0662191.	
19-OCT-2000; 2000US-0693036.	
29-NOV-2000; 2000US-0727344.	
(HYSE-) HYSEQ INC.	
Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;	
Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;	
Zhao QA, Zhou P, Goodrich R, Drmanac RT;	
WPI; 2001-442253/47.	
N-PSDB; AAI59160.	
Novel nucleic acids and polypeptides, useful for treating disorders	
such as central nervous system injuries -	
Example 4; SEQ ID NO 3149; 10076pp; English.	

XX	The invention relates to human nucleic acids (AA157798-AA161369) and
CC	the encoded polypeptides (AAM38642-AAM42213) with neurotrophic,
CC	In immunosuppressant and cytoskeletal activity. The polynucleotides are useful
CC	in gene therapy. A composition containing a polypeptide or polynucleotide
CC	of the invention may be used to treat diseases of the peripheral nervous
CC	system, such as peripheral nervous injuries, peripheral neuropathy and
CC	localised neuropathies and central nervous system diseases, such as
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC	utilisation of the activities such as: Immune system suppression,
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC	assays for receptor activity, arthritis and inflammation, leukaemias and
CC	C.N.S disorders.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification.
CC	
XX	Sequence 602 AA:
SQ	
Query Match	Best Local Similarity 100.0% Score 33; DB 22; Length 602;
Matches 6;	Conservative 100.0%; Pred. NO. 8.1e+02;
Mismatches 0;	Gaps 0;
Indels 0;	
DY	1 GCGTSG 6 
Db	83 ggtsg 88
RESULT 30	
AAM40003	
ID	AAM40003 standard; Protein: 625 AA.
XX	AAM40003;
DT	22-OCT-2001 (first entry)
DE	Human polypeptide SEQ ID NO 3148.
XX	
XX	Human; neurotropic; immunosuppressant; cytoskeletal; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; chemokine; thrombolytic; drug screening; arthritis; inflammation; leukemia.
OS	Homo sapiens.
PN	WO200153312-A1.
PD	26-JUL-2001.
PF	26-DEC-2000; 2000WO-US34263.
PR	21-JAN-2000; 2000US-0488725. 23-APR-2000; 2000US-052317. 09-JUL-2000; 2000US-0598042. 19-JUL-2000; 2000US-0620312. 03-AUG-2000; 2000US-0634450. 14-SEP-2000; 2000US-0662191. 19-OCT-2000; 2000US-0693036. 29-NOV-2000; 2000US-0727344.
PA	(HYSE-) HYSEQ INC.
PI	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D; Wang J, Wang Z, Wehtman T, Xu C, Xue AJ, Yang Y, Zhang J; Zhao QA, Zhou P, Goodrich R, Drmanac RJ.
DR	WI: 2001-442253/47. N-PDB: AA159159.

PT Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -  
XX  
PS Example 4; SEQ ID NO 3148; 10078bp; English.  
XX  
CC The invention relates to human nucleic acids (AA157798-AA161369) and  
CC the encoded polypeptides (AAM38642-AAM42213) with neurotropic,  
CC immunosuppressant and cyostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localized neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
CC  
SQ Sequence 625 AA;  
XX

Query Match 100.0%; Score 33; DB 22; Length 625;  
Best Local Similarity 100.0%; Pred. No. 8.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 GCGTSG 6  
|||||  
DB 106 99gtsq 111

RESULT 31  
AAB94603  
ID AAB94603 standard; Protein; 678 AA.  
XX  
AC AAB94603;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human protein sequence SEQ ID NO:15428.  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX  
OS Homo sapiens.  
XX  
PN EP1074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
PF 28-JUL-2000; 2000EP-0116126.  
XX  
PR 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
DR WPI; 2001-318749/34.  
XX  
PT Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
PS Claim 8; SEQ ID 15428; 2537pp + CD ROM; English.

XX  
CC The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
CC  
SQ Sequence 678 AA;  
XX

Query Match 100.0%; Score 33; DB 22; Length 678;  
Best Local Similarity 100.0%; Pred. No. 9e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 GCGTSG 6  
|||||  
DB 659 99gtsq 664

RESULT 32  
AAY84686  
ID AAY84686 standard; Protein; 2071 AA.  
XX  
AC AAY84686;  
XX  
DT 08-AUG-2000 (first entry)  
XX  
DE Amino acid sequence of AG003, a GTPase activating protein.  
XX  
KW GTP-binding protein; AG003; GTPase activating protein; fungal growth;  
KW phosphatidylinositol-4 kinase; cytokinesis gene; fungal development;  
KW fungicidal; filamentous fungi; plant pathogen; Septoria tritici;  
KW Stagonospora nodorum; Magnaporthe grisea; human pathogen;  
KW Candida albicans; Aspergillus fumigatus.  
XX  
XX  
OS Ashbya gossypii.  
XX  
PN WO200022133-A1.  
XX  
PD 20-APR-2000.  
XX  
PF 06-OCT-1999; 99WO-EP07501.  
XX  
PR 08-OCT-1998; 98US-0168804.  
XX  
XX  
PA (NOVS ) NOVARTIS AG.  
PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GRS MBH.  
XX  
XX  
PI Gaffney TD, Wendland J, Dietrich F, Philippsen P, Goff SA;  
XX  
DR WPI; 2000-317988/27.  
DR N-PSDB; AAA14503.  
XX  
PT Ashbya gossypii nucleotide and protein sequences useful for identifying  
PT fungal growth inhibitors -  
XX

PS Claim 2; Page 48-60; 93pp; English.

XX The present sequence represents a putative GTPase activating protein,

CC designated AG003. The specification also describes putative GTP

CC binding protein genes, putative phosphatidylinositol-4 kinase

CC protein gene, and a putative cytokinesis gene. These genes are

CC essential for fungal growth and development. The proteins can be used

CC in methods to identify compounds that have fungicidal activity. Compounds

CC with fungicidal activity can be used for suppressing fungal growth,

CC especially of filamentous fungi. Fungi that can be suppressed include

CC plant pathogens (e.g. *Septoria tritici*, *Stagonospora nodorum*, and

CC *Magnaporthe oryzae*) and human pathogens (e.g. *Candida albicans*, and

CC *Aspergillus fumigatus*).

XX

SQ Sequence 2071 AA;

Query Match 100.0%; Score 33; DB 21; Length 2071;

Best Local Similarity 100.0%; Pred. No. 2.4e+03;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6

111111

DB 199 ggtstg 204

RESULT 33

AAB86196

ID AAB86196 standard; Protein: 3025 AA.

XX

AC AAB86196;

XX

DT 21-AUG-2001 (first entry)

XX

DE HIV-1 subtype C protein fragment #2.

XX

KM Infection; diagnosis: human; humoral immune response; antiviral;

KW cellular immune response; vaccine; treatment; gene therapy.

XX

OS Human immunodeficiency virus type 1.

XX

FH Key Location/Qualifiers

FT MISC-difference 1..3025

FT /label= "xaa

FT /note= "xaa represents a stop codon"

XX

PN DE10056747-A1.

XX

PD 31-MAY-2001.

XX

PF 16-NOV-2000; 2000DE-1056747.

XX

PR 16-NOV-1999; 99DE-1055089.

XX

PA (SHAQ/) SHAO Y.

PA (GENE-) GENEA RT GMBH GES ANGEWANDTE BIOTECHNOLOG.

XX

PI Wagner R, Wolf H, Shao Y, Graf M;

XX

DR WPI: 2001-336417/36.

XX

PT New nucleic acid sequences from a human immune deficiency virus

PT Interubtype, useful for treatment, prevention and diagnosis of

PT infection

XX

PS Disclosure; Fig 8A-O; 48pp; German.

XX

CC This invention describes a novel polynucleotide isolated from human

CC immunodeficiency virus type 1 subtype C/B' which can be used for the

CC induction of specific humoral and cellular immune responses. (1) and

CC polypeptides (II) encoded by them, are useful in pharmaceuticals,

CC vaccines and diagnostic agents, particularly for treatment or prevention

CC of human immune deficiency virus-1 (HIV-1) infections, also for rational

CC design of test or therapeutic reagents, or gene therapy vectors.

CC Polypeptides, especially antibodies, specifically directed against (II)

CC are similarly useful as pharmaceutical and diagnostic agents. (I) are

CC specific for intersubtype C/B' of HIV-1 so are useful in regions

CC (particularly China and South-East Asia) where this subtype is prevalent.

CC The products of the invention have antiviral activity. This sequence

CC represents a protein encoded by the HIV-1 subtype C genome described in

CC the method of the invention.

XX

SQ Sequence 3025 AA;

Query Match 100.0%; Score 33; DB 22; Length 3025;

Best Local Similarity 100.0%; Pred. No. 3.4e+03;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6

111111

DB 2639 ggtstg 2644

RESULT 34

AAW90042

ID AAW90042 standard; Protein: 32 AA.

XX

AC AAW90042;

XX

DT 26-FEB-1999 (first entry)

XX

DE B. steatothermophilus alanine dehydrogenase fragment #1.

XX

KM Phenylalanine dehydrogenase; pdh; L-amino acid; phenylpyruvic acid;

KW aspartame production; phenylketonuria; infant; alanine dehydrogenase;

KW diet.

XX

OS Bacillus steatothermophilus.

XX

PN US5851810-A.

XX

PD 22-DEC-1998.

XX

PF 05-JUN-1995; 95US-0461990.

XX

PR 05-JUN-1995; 95US-0461990.

XX

PA (YESH ) UNIV YESHIVA EINSTEIN COLLEGE.

XX

PI Blanchard JS;

XX

DR WPI: 1999-080406/07.

XX

PT DNA encoding Rhodococcus L-phenylalanine dehydrogenase - for

PT production of recombinant enzyme

XX

PS Disclosure; Fig 4; 30pp; English.

XX

CC This sequence is used to describe a method which results in the isolation

CC of a phenylalanine dehydrogenase protein, pdh, from *Rhodococcus* sp. M4.

CC The encoding pdh nucleic acid is used to produce recombinant *Rhodococcus*

CC L-phenylalanine dehydrogenase, which can be used to produce L-amino acids

CC from the corresponding alpha-keto acids, e.g. to produce L-phenylalanine

CC from phenylpyruvic acid (e.g. for the production of aspartame), to screen

CC for phenylketonuria in newborn infants and to monitor the efficacy of

CC low-phenylalanine diets.

XX

SQ Sequence 32 AA;

Query Match 90.9%; Score 30; DB 20; Length 32;

Best Local Similarity 83.3%; Pred. No. 1.8e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6



Db 6 999tag 11

RESULT 35

AAW77518  
ID AAW77518 standard; Protein; 66 AA.

AC AAW77518;

DT 30-OCT-1998 (first entry)

DE Staphylococcus aureus protein of unknown function.

XX Staphylococcus aureus protein; immune response induction; eye infection;  
KW antibody production; T-cell immune response; gastrointestinal infection;  
KW respiratory infection; inhibitor; bacterial infection; cardiac infection;  
KW central nervous system; kidney infection; urinary tract infection;  
KW antimicrobial compound identification; broad spectrum antibiotic;  
therapy.

OS Staphylococcus aureus.

FT Key Location/Qualifiers

FT Misc-difference 1..66 /note- "residues designated X are unspecified, and  
represented as Xaa in the specification"

EP841394-A2.

13-MAY-1998.

24-SEP-1997; 97EP-0307485.

24-SEP-1996; 96US-0027032.

(SMRK ) SMITHKLINE BEECHAM CORP.  
(SMRK ) SMITHKLINE BEECHAM PLC.

PI Black MT, Burnham MKR, Hodgson JE, Knowles DJC;  
PI Lonetto MA, Nicholas RO, Pratt JM, Reichard RW, Rosenberg M;  
PI Ward JM;

DR WPI: 1998-252940/23.

DR N-PSDB: AAW53319.

XX New nucleic acid sequences from Staphylococcus aureus WCH029 -  
PT useful in vaccines and for treatment of bacterial infections of e.g.  
PT respiratory tract and central nervous system

PS Claim 11; Page 228; 390pp; English.

XX This sequence represents a Staphylococcus aureus protein of unknown  
CC function, and is encoded by a DNA sequence of the invention.  
CC The DNA sequences were isolated from Staphylococcus aureus WCH029  
CC (NCMB 40771). Host cells containing the DNA sequences are used to  
CC produce polypeptides or fragments. The proteins are used in the treatment  
CC of disease, for inducing an immune response by administering them, to  
CC produce antibody and/or T-cell immune response. Antagonists of the  
CC proteins are used for the inhibition of bacterial polypeptides.  
CC Conditions which may be treated include bacterial infections, especially  
CC respiratory, cardiac, gastrointestinal, central nervous, eye, kidney,  
CC urinary tract, skin, bones and joints. The proteins can also be used to  
CC identify antimicrobial compounds which are broad spectrum antibiotics,  
CC especially useful in the treatment of H. pylori infection.

XX Sequence 66 AA;

Query Match 90.9%; Score 30; DB 19; Length 66;  
Best Local Similarity 83.3%; Pred. NO. 3.4e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCTSG 6  
Db 14 999tag 19

RESULT 36  
AAC15636  
ID AAC15636 standard; Protein; 81 AA.

AC AAC15636;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 15966.

XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

25-FEB-1999; 99US-0121825.

05-MAR-1999; 99US-0123180.

09-MAR-1999; 99US-0123548.

23-MAR-1999; 99US-0125788.

25-MAR-1999; 99US-0126264.

29-MAR-1999; 99US-0126785.

01-APR-1999; 99US-0127462.

06-APR-1999; 99US-0128234.

08-APR-1999; 99US-0128714.

16-APR-1999; 99US-0129845.

19-APR-1999; 99US-0130077.

23-APR-1999; 99US-0130449.

28-APR-1999; 99US-0130891.

30-APR-1999; 99US-0131449.

30-APR-1999; 99US-0132048.

04-MAY-1999; 99US-0132407.

05-MAY-1999; 99US-0132484.

06-MAY-1999; 99US-0132486.

07-MAY-1999; 99US-0132487.

11-MAY-1999; 99US-0132863.

14-MAY-1999; 99US-0134256.

14-MAY-1999; 99US-0134218.

14-MAY-1999; 99US-0134219.

14-MAY-1999; 99US-0134221.

14-MAY-1999; 99US-0134370.

18-MAY-1999; 99US-0134768.

19-MAY-1999; 99US-0134941.

20-MAY-1999; 99US-0135124.

21-MAY-1999; 99US-0135353.

24-MAY-1999; 99US-0135629.

25-MAY-1999; 99US-0136021.

27-MAY-1999; 99US-0136392.

28-MAY-1999; 99US-0136782.

01-JUN-1999; 99US-0137222.

03-JUN-1999; 99US-0137528.

04-JUN-1999; 99US-0137502.

07-JUN-1999; 99US-0137724.

08-JUN-1999; 99US-0138094.

10-JUN-1999; 99US-0138540.

10-JUN-1999; 99US-0138847.

14-JUN-1999; 99US-0139119.

16-JUN-1999; 99US-0139452.

16-JUN-1999; 99US-0139453.

17-JUN-1999; 99US-0139492.

```
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 23-JUN-1999; 99US-0140685.
PR 24-JUN-1999; 99US-0140685.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145291.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.

PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149475.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150866.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160747.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 90.9%; Score 30; DB 21; Length 81;
Best Local Similarity 83.3%; Pred. No. 4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Caps 0;
```

RESULT 37  
ANG15635

ID	AAAG15635 standard; Protein; 83 AA.	PR	18-JUN-1999;	99US-0139461
XX		PR	18-JUN-1999;	99US-0139462
XX	AAAG15635;	PR	18-JUN-1999;	99US-0139463
XX		PR	18-JUN-1999;	99US-0139464
DT	17-OCT-2000 (first entry)	PR	18-JUN-1999;	99US-0139465
XX		PR	18-JUN-1999;	99US-0139466
XX	Arabidopsis thaliana protein fragment SEQ ID NO: 15965.	PR	21-JUN-1999;	99US-0139467
XX		PR	22-JUN-1999;	99US-0139468
XX		PR	23-JUN-1999;	99US-0140353
KW	Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.	PR	23-JUN-1999;	99US-0140354
KW		PR	24-JUN-1999;	99US-0140355
OS	Arabidopsis thaliana.	PR	28-JUN-1999;	99US-0140693
XX		PR	29-JUN-1999;	99US-0140694
PN	EP1033405-A2.	PR	30-JUN-1999;	99US-0141287
PD		PR	01-JUL-1999;	99US-0141842
XX		PR	01-JUL-1999;	99US-0142154
XX		PR	02-JUL-1999;	99US-0143055
XX		PR	06-JUL-1999;	99US-0143390
XX		PR	08-JUL-1999;	99US-0142803
PF	25-FEB-2000; 2000EP-0301439.	PR	09-JUL-1999;	99US-0145920
XX		PR	12-JUL-1999;	99US-0142977
XX		PR	13-JUL-1999;	99US-0143542
PR	25-FEB-1999;	PR	15-JUL-1999;	99US-0144085
PR	05-MAR-1999;	PR	16-JUL-1999;	99US-0144086
PR	09-MAR-1999;	PR	19-JUL-1999;	99US-0144331
PR	23-MAR-1999;	PR	19-JUL-1999;	99US-0144332
PR	29-MAR-1999;	PR	19-JUL-1999;	99US-0144333
PR	01-APR-1999;	PR	19-JUL-1999;	99US-0144334
PR	06-APR-1999;	PR	19-JUL-1999;	99US-0144335
PR	08-APR-1999;	PR	20-JUL-1999;	99US-0144336
PR	16-APR-1999;	PR	20-JUL-1999;	99US-0144337
PR	19-APR-1999;	PR	20-JUL-1999;	99US-0144338
PR	21-APR-1999;	PR	20-JUL-1999;	99US-0144339
PR	23-APR-1999;	PR	21-JUL-1999;	99US-0144340
PR	28-APR-1999;	PR	21-JUL-1999;	99US-0144341
PR	30-APR-1999;	PR	21-JUL-1999;	99US-0145086
PR	30-APR-1999;	PR	22-JUL-1999;	99US-0145087
PR	04-MAY-1999;	PR	22-JUL-1999;	99US-0145088
PR	05-MAY-1999;	PR	22-JUL-1999;	99US-0145089
PR	06-MAY-1999;	PR	22-JUL-1999;	99US-0145192
PR	07-MAY-1999;	PR	23-JUL-1999;	99US-0145145
PR	11-MAY-1999;	PR	23-JUL-1999;	99US-0145218
PR	14-MAY-1999;	PR	23-JUL-1999;	99US-0145222
PR	14-MAY-1999;	PR	26-JUL-1999;	99US-0145276
PR	14-MAY-1999;	PR	27-JUL-1999;	99US-0145593
PR	14-MAY-1999;	PR	27-JUL-1999;	99US-0145598
PR	14-MAY-1999;	PR	27-JUL-1999;	99US-0145599
PR	18-MAY-1999;	PR	28-JUL-1999;	99US-0145919
PR	19-MAY-1999;	PR	28-JUL-1999;	99US-0146386
PR	20-MAY-1999;	PR	02-AUG-1999;	99US-0146388
PR	21-MAY-1999;	PR	02-AUG-1999;	99US-0146389
PR	24-MAY-1999;	PR	03-AUG-1999;	99US-0146389
PR	25-MAY-1999;	PR	04-AUG-1999;	99US-0147204
PR	27-MAY-1999;	PR	05-AUG-1999;	99US-0147302
PR	28-MAY-1999;	PR	05-AUG-1999;	99US-0147303
PR	01-JUN-1999;	PR	06-AUG-1999;	99US-0147303
PR	03-JUN-1999;	PR	06-AUG-1999;	99US-0147416
PR	04-JUN-1999;	PR	09-AUG-1999;	99US-0147493
PR	07-JUN-1999;	PR	10-AUG-1999;	99US-0147935
PR	08-JUN-1999;	PR	11-AUG-1999;	99US-0148171
PR	10-JUN-1999;	PR	11-AUG-1999;	99US-0148319
PR	14-JUN-1999;	PR	12-AUG-1999;	99US-0148341
PR	16-JUN-1999;	PR	13-AUG-1999;	99US-0148565
PR	17-JUN-1999;	PR	13-AUG-1999;	99US-0148684
PR	18-JUN-1999;	PR	16-AUG-1999;	99US-0149368
PR	18-JUN-1999;	PR	17-AUG-1999;	99US-0149368
PR	18-JUN-1999;	PR	18-AUG-1999;	99US-0149426
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149723
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149723
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149929
PR	18-JUN-1999;	PR	23-AUG-1999;	99US-0149929

```
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154079.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159285.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160960.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 90.9%; Score 30; DB 21; Length 83;  
Best Local Similarity 83.3%; Pred. No. 4.1e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGGTSG 6  
||||:|  
Db 37 gggtag 42

RESULT 38  
AAW90054  
ID AAW90054 standard; Protein; 95 AA.

XX AC AAW90054;

XX DT 26-FEB-1999 (first entry)

DE B. steaerotherophilus alanine dehydrogenase fragment #2.

```
XX XX  
KW phenylalanine dehydrogenase; pdh; L-amino acid; phenylpyruvic acid;  
KW aspartame production; phenylketonuria; infant; alanine dehydrogenase;  
KW diet.  
XX  
OS Bacillus steaerotherophilus.  
XX  
PN US5851810-A.  
XX  
PD 22-DEC-1998.  
XX  
PF 05-JUN-1995; 95US-0461990.  
XX  
PR 05-JUN-1995; 95US-0461990.  
XX  
PA (YESH ) UNIV YESHIVA EINSTEIN COLLEGE.  
XX  
PI Blanchard JS;  
XX  
DR WPI; 1999-080406/07.  
XX  
PT DNA encoding Rhodococcus L-phenylalanine dehydrogenase - for  
PT production of recombinant enzyme  
XX  
PS Disclosure: Fig 5; 30pp; English.  
XX  
CC This sequence is used to describe a method which results in the isolation  
CC of a phenylalanine dehydrogenase protein, pdh, from Rhodococcus sp. M4.  
CC The encoding pdh nucleic acid is used to produce recombinant Rhodococcus  
CC L-phenylalanine dehydrogenase, which can be used to produce L-amino acids  
CC from the corresponding alpha-keto acids, e.g. to produce L-phenylalanine  
CC from phenylpyruvic acid (e.g. for the production of aspartame), to screen  
CC for phenylketonuria in newborn infants and to monitor the efficacy of  
CC low-phenylalanine diets.  
XX  
SQ Sequence 95 AA:
```

Query Match 90.9%; Score 30; DB 20; Length 95;  
Best Local Similarity 83.3%; Pred. No. 4.6e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGGTSG 6  
||||:|  
Db 67 gggtag 72

RESULT 39

AAW27159  
ID AAW27159 standard; Protein; 116 AA.

XX AC AAW27159;

XX DT 17-OCT-2000 (first entry)

DE Zea mays protein fragment SEQ ID NO: 31888.

KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence; corn.

XX OS Zea mays subsp. mays.

XX OS EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

XX PR 05-MAR-1999; 99US-0123180.

XX PR 09-MAR-1999; 99US-0123548.

XX PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134236.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151530.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.

PR 08-OCT-1999; 99US-0158222.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 25-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 90.98; Score 30; DB 21; Length 116;  
Best Local Similarity 83.38; Pred. No. 5; Se+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GCGTSG 6  
Db 37 g99tag 42

## RESULT 40

AAC06217 standard; Protein; 256 AA.

AAC06217;

17-OCT-2000 (first entry)

Arabidopsis thaliana protein fragment SEQ ID NO: 2910.

XX Protein Identification: signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.

OS Arabidopsis thaliana.

XX EPI033405-A2.

PN 06-SEP-2000.

PD 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 23-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140655.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0141847.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 28-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 04-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 09-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 10-AUG-1999; 99US-0147935.  
PR 11-AUG-1999; 99US-0148171.  
PR 12-AUG-1999; 99US-0148319.  
PR 13-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158363.  
PR 13-OCT-1999; 99US-0158293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.

PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 90.9%; Score 30; DB 21; Length 256;  
Best Local Similarity 83.3%; Pred. No. 11e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
OY 1 GGGTSG 6  
Db 231 gggttg 236

Search completed: February 4, 2002, 08:03:03  
Job time: 167 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 4, 2002, 08:00:16 ; Search time 12.9 Seconds  
(without alignments)  
17.053 Million cell updates/sec

Title: us-09-642-660-10  
Perfect score: 33  
Sequence: 1 GGCTSG 6

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database : SwissProt\_39:\*  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33	100.0	102	1	VNI2_MERTL
2	33	100.0	284	1	CDX4_HUMAN
3	33	100.0	559	1	MDL1_PRODU
4	33	100.0	563	1	MDL1_PROSE
5	33	100.0	568	1	PTLB_LACIA
6	33	100.0	568	1	PTLB_STRMU
7	33	100.0	572	1	PTLB_STAAU
8	33	100.0	577	1	PTLB_LACCA
9	33	100.0	625	1	GCKR_HUMAN
10	33	100.0	626	1	GCKR_HUMAN
11	33	100.0	629	1	K2C3_HUMAN
12	33	100.0	668	1	UL52_HCMVA
13	33	100.0	677	1	SP87_DICDI
14	33	100.0	780	1	NH48_CAEBL
15	33	100.0	796	1	KF3C_RAT
16	33	100.0	923	1	CLAB_RAT
17	33	100.0	926	1	CLAB_LYCES
18	33	100.0	932	1	CLAB_LYCES
19	33	100.0	936	1	LY55_CAEBL
20	33	100.0	936	1	NRDI_ECOLI
21	33	100.0	936	1	YJF5_YEAST
22	33	100.0	936	1	YJF5_YEAST
23	33	100.0	936	1	YJF5_YEAST
24	33	100.0	936	1	YJF5_YEAST
25	33	100.0	936	1	YJF5_YEAST
26	33	100.0	936	1	YJF5_YEAST
27	33	100.0	936	1	YJF5_YEAST
28	33	100.0	936	1	YJF5_YEAST
29	33	100.0	936	1	YJF5_YEAST
30	33	100.0	936	1	YJF5_YEAST
31	33	100.0	936	1	YJF5_YEAST
32	33	100.0	936	1	YJF5_YEAST
33	33	100.0	936	1	YJF5_YEAST

Result	Score	Query Match	Length	ID	Description
34	30	90.9	410	1	MURG_MYCTU
35	30	90.9	463	1	YAB8_MYCTU
36	30	90.9	573	1	MDL2_PROSE
37	30	90.9	576	1	MDL2_PROSE
38	30	90.9	608	1	OM70_HUMAN
39	30	90.9	611	1	OM70_HUMAN
40	30	90.9	916	1	YJRA_THEMA
41	30	90.9	3828	1	TRX_DROVI
42	29	87.9	33	1	ALOX_PICPA
43	29	87.9	123	1	BCCP_PROPR
44	29	87.9	135	1	YF84_MYCPN
45	29	87.9	166	1	K2C5_BOVIN
46	29	87.9	226	1	SPC1_CANFA
47	29	87.9	278	1	T2D7_DROME
48	29	87.9	293	1	MHPC_ECOLI
49	29	87.9	306	1	HMUX_DROPS
50	29	87.9	313	1	YCGG_ECOLI
51	29	87.9	316	1	LORI_HUMAN
52	29	87.9	330	1	HX11_HUMAN
53	29	87.9	330	1	HXCD_HUMAN
54	29	87.9	338	1	GRP_ARATH
55	29	87.9	338	1	HXDB_HUMAN
56	29	87.9	339	1	MURG_THEMA
57	29	87.9	344	1	HM10_CAEBL
58	29	87.9	344	1	JUNB_MOUSE
59	29	87.9	344	1	JUNB_MOUSE
60	29	87.9	344	1	JUNB_MOUSE
61	29	87.9	344	1	JUNB_MOUSE
62	29	87.9	344	1	JUNB_MOUSE
63	29	87.9	344	1	JUNB_MOUSE
64	29	87.9	344	1	JUNB_MOUSE
65	29	87.9	344	1	JUNB_MOUSE
66	29	87.9	344	1	JUNB_MOUSE
67	29	87.9	344	1	JUNB_MOUSE
68	29	87.9	344	1	JUNB_MOUSE
69	29	87.9	344	1	JUNB_MOUSE
70	29	87.9	344	1	JUNB_MOUSE
71	29	87.9	344	1	JUNB_MOUSE
72	29	87.9	344	1	JUNB_MOUSE
73	29	87.9	344	1	JUNB_MOUSE
74	29	87.9	344	1	JUNB_MOUSE
75	29	87.9	344	1	JUNB_MOUSE
76	29	87.9	344	1	JUNB_MOUSE
77	29	87.9	344	1	JUNB_MOUSE
78	29	87.9	344	1	JUNB_MOUSE
79	29	87.9	344	1	JUNB_MOUSE
80	29	87.9	344	1	JUNB_MOUSE
81	29	87.9	344	1	JUNB_MOUSE
82	29	87.9	344	1	JUNB_MOUSE
83	29	87.9	344	1	JUNB_MOUSE
84	29	87.9	344	1	JUNB_MOUSE
85	29	87.9	344	1	JUNB_MOUSE
86	29	87.9	344	1	JUNB_MOUSE
87	29	87.9	344	1	JUNB_MOUSE
88	29	87.9	344	1	JUNB_MOUSE
89	29	87.9	344	1	JUNB_MOUSE
90	29	87.9	344	1	JUNB_MOUSE
91	29	87.9	344	1	JUNB_MOUSE
92	29	87.9	344	1	JUNB_MOUSE
93	29	87.9	344	1	JUNB_MOUSE
94	29	87.9	344	1	JUNB_MOUSE
95	29	87.9	344	1	JUNB_MOUSE
96	29	87.9	344	1	JUNB_MOUSE
97	29	87.9	344	1	JUNB_MOUSE
98	29	87.9	344	1	JUNB_MOUSE
99	29	87.9	344	1	JUNB_MOUSE
100	29	87.9	344	1	JUNB_MOUSE

## ALIGNMENTS

```

YN12_METTLL
ID YN12_METTLL STANDARD; PRT: 102 AA.
AC P05410;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE HYPOTHEICAL PROTEIN IN NIFH2 3'REGION (FRAGMENT).
OS Methanococcus thermolithotrophicus.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2186;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88259240; PubMed=2838639;
RA Soullard N., Magot M., Posset O., Sibold L.;
RT "Nucleotide sequence of regions homologous to nifH (nitrogenase Fe
protein) from the nitrogen-fixing archaebacteria Methanococcus
thermolithotrophicus and Methanobacterium Ivanovii: evolutionary
implications."
RL J. Mol. Evol. 27:65-76(1988).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X07500; CAA30382.1; -
DR PIR: S00739; S00739.
DR InterPro: IPR000205; NAD_binding.
KW Hypothetical protein; Nitrogen fixation.
FT NON_TER 102 102
SQ SEQUENCE 102 AA; 11362 MW; 140266961975FE2 CRC64;

Query Match
Best Local Similarity 100.0%; Score 33; DB 1; Length 102;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGTSG 6
Db 7 GGGTSG 12

RESULT 2
CDX4_HUMAN STANDARD; PRT: 284 AA.
AC O14627;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HOMEBOX PROTEIN CDX-4 (CAUDAL-TYPE HOMEBOX PROTEIN 4).
GN CDX4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eultheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Chen E.Y., States D.J., Mazarrella R.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RT [2]
RP SEQUENCE FROM N.A.
RA Chlaroni P., Colleaux L., Briault S., Losel A.M., Moraine C.,
RA Fontes M.;
RT "Genomic structure of the human CDX4 gene: a potential candidate for
the FG syndrome."
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CAUDAL FAMILY OF HOMEBOX PROTEINS.
CC -----

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF003530; AAB6319.1; -
DR EMBL: AF029879; AAD01894.1; -
DR EMBL: AF029877; AAD01894.1; JOINED.
DR EMBL: AF029878; AAD01894.1; JOINED.
DR HSSP: P02835; 1FTZ.
DR MIM: 300025; -
DR InterPro: IPR001827; Antennapedia.
DR InterPro: IPR000047; HTH_repressor.
DR InterPro: IPR001356; Homeobox.
DR Pfam: PF00046; homeobox; 1.
DR PRINTS: PR00024; HOMEBOX.
DR PRINTS: PR00025; ANTENNAPEDIA.
DR PRINTS: PR00031; HTHREPRESSR.
DR SMART: SM00389; HOX; 1.
DR PROSITE: PS00027; HOMEBOX_1; 1.
DR PROSITE: PS00071; HOMEBOX_2; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;
KW Transcription regulation.
FT DNA_BIND 173 232
SQ SEQUENCE 284 AA; 30480 MW; 96D365D9F5F6D6D7 CRC64;

```

```

Query Match
Best Local Similarity 100.0%; Score 33; DB 1; Length 284;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGTSG 6
Db 120 GGGTSG 125

RESULT 3
MDL1_PRUDU STANDARD; PRT: 559 AA.
AC O24243;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE (R)-MANDELONITRILE LYASE ISOFORM 1 PRECURSOR (EC 4.1.2.10)
DE (HYDROXYNITRILE LYASE 1) ((R)-OXYNITRILEASE 1).
GN MDL1.
OS Prunus dulcis (Almond) (Prunus amygdalus).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids I; Rosales; Rosaceae; Prunus;
OX NCBI_TaxID=3755;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-CV, TEXAS; TISSUE=Flower;
RA Snulves M.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN CYANOGENIN, THE RELEASE OF HCN FROM INJURED
CC TISSUES. CATALYSES THE STEREOSPECIFIC ADDITION OF HCN TO A VARIETY
CC OF ALDEHYDES IN VITRO. IS A MAJOR SEED CONSTITUENT, AND COULD HAVE
CC THE ADDITIONAL ROLE OF A STORAGE FORM FOR REDUCED NITROGEN (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: MANDELONITRILE - CYANIDE + BENZALDEHYDE.
CC -1- COFACTOR: FAD (BY SIMILARITY).
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE GMC OXIDOREDUCTASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```



```

RL J. Biol.Chem. 265:22554-22560(1990).
CC -1- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC -TRANSFERT SYSTEM. THE IID DOMAINS CONTAIN THE SUGAR BINDING SITE
CC AND THE TRANSMEMBRANE CHANNEL. THE IIA DOMAIN CONTAINS THE PRIMARY
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC THE SUGAR.
CC
CC -1- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHISTIDINE + SUGAR =
CC PROTEIN HISTIDINE + SUGAR PHOSPHATE.
CC
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC
CC -1- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
CC
CC -1- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
CC
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
CC
CC DR EMBL: M60447; AAA25182.1; -.
CC DR PIR: B23696; B23696.
CC DR InterPro: IPR003352; PTS_EIIC.
CC DR InterPro: IPR003501; PTS_IIB.
CC DR Pfam: PF02378; PTS_EIIC; 1.
CC DR Pfam: PF02302; PTS_IIB; 1.
CC
CC KW Phosphotransferase system; Sugar transport; Transferase;
CC Phosphorylation; Transmembrane; Plasmaid.
CC
CC FT DOMAIN 1 568 ? EIIC DOMAIN.
CC FT TRANSSEM 30 50 EIIB DOMAIN.
CC FT TRANSSEM 62 82 POTENTIAL.
CC FT TRANSSEM 103 123 POTENTIAL.
CC FT TRANSSEM 128 148 POTENTIAL.
CC FT TRANSSEM 183 203 POTENTIAL.
CC FT TRANSSEM 222 242 POTENTIAL.
CC FT TRANSSEM 246 266 POTENTIAL.
CC FT TRANSSEM 283 303 POTENTIAL.
CC FT TRANSSEM 316 336 POTENTIAL.
CC FT TRANSSEM 339 359 POTENTIAL.
CC FT TRANSSEM 367 387 POTENTIAL.
CC FT TRANSSEM 390 410 POTENTIAL.
CC FT TRANSSEM 466 486 POTENTIAL.
CC
CC SO SEQUENCE 568 AA; 61563 MW; BBE5CA7736762446 CRC64;

```

```

RP SEQUENCE FROM N.A.
RC STRAIN-0A130 / SEROTYPE C;
RX MEDLINE-93015655; PubMed-1400164;
RA Rosey E.L., Stewart G.C.;
RT "Nucleotide and deduced amino acid sequences of the lacR, lacABCD,
RT and lacE genes encoding the repressor, tagatose 6-phosphate gene
RT cluster, and sugar-specific phosphotransferase system components of
RT the lactose operon of streptococcus mutans.";
RL J. Bacteriol. 174:6159-6170(1992).
RN [2]
RC SEQUENCE FROM N.A.
RP STRAIN-0A130 / SEROTYPE C;
RX MEDLINE-94103744; PubMed-8277252;
RA Honeyman A.L., Curtiss R. III;
RT "Isolation, characterization and nucleotide sequence of the
RT Streptococcus mutans lactose-specific enzyme II (lacE) gene of the
RT PTS and the phospho-beta-galactosidase (lacG) gene.";
RL J. gen. Microbiol. 139:2685-2694(1993).
RC -1- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC AND THE TRANSMEMBRANE CHANNEL. THE IIA DOMAIN CONTAINS THE PRIMARY
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSHO-HPR); IIA TRANSFERS ITS
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC THE SUGAR.
CC -1- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHISTIDINE + SUGAR =
CC PROTEIN HISTIDINE + SUGAR PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
CC -1- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
CC -----
CC THE SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, L18993; AAA16449.1; -
DR InterPro: IPR003352; PTS_EIIC.
DR InterPro: IPR003501; PTS_IIB.
DR Pfam: PF02378; PTS_EIIC.1.
DR Pfam: PF02302; PTS_IIB.1.
KM Phosphotransferase system; Sugar transport; Transferase;
KM Phosphorylation; Transmembrane.
KW DOMAIN 1 ?
FT DOMAIN ? 568
FT TRANSMEM 30 50 EIIB DOMAIN.
FT TRANSMEM 65 85 POTENTIAL.
FT TRANSMEM 103 123 POTENTIAL.
FT TRANSMEM 128 148 POTENTIAL.
FT TRANSMEM 183 203 POTENTIAL.
FT TRANSMEM 223 243 POTENTIAL.
FT TRANSMEM 246 266 POTENTIAL.
FT TRANSMEM 283 303 POTENTIAL.
FT TRANSMEM 316 336 POTENTIAL.
FT TRANSMEM 339 359 POTENTIAL.
FT TRANSMEM 381 401 POTENTIAL.
FT TRANSMEM 465 485 POTENTIAL.
SQ SEQUENCE 568 AA; 61419 MW; A129FDD14282EAAE CRC64;

```

```

PTLB_STAU          STANDARD:      PRT:      572 AA.
ID PTLB_STAU
AC P1162:
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE PTS SYSTEM, LACTOSE-SPECIFIC IIBC COMPONENT (EIIIC-LAC) (LACTOSE-
DE PERASE IIBC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, BC COMPONENT)
DE (EC 2.7.1.69) (EII-LAC).
GN LACE.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxId=1280;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8059023; PubMed=2824493;
RA Breidt F. Jr., Hengstenberg W., Finkeldei U., Stewart G.C.:
RT Identification of the genes for the lactose-specific components of
RT the phosphotransferase system in the lac operon of Staphylococcus
RT aureus."
RT J. Biol. Chem. 262:16444-16449(1987).
CC -1- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
CC AND THE TRANSMEMBRANE CHANNEL. THE IIA DOMAIN CONTAINS THE PRIMARY
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC THE SUGAR.
CC -1- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHISTIDINE + SUGAR -
CC PROTEIN HISTIDINE + SUGAR PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: CONTAINS A PTS EIIIB DOMAIN.
CC -1- SIMILARITY: CONTAINS A PTS EIIIC DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: J03479; AAA26649.1; -.
CC PIR: B28474; B28474.
CC InterPro: IPR003352; PTS_EIIC.
CC InterPro: IPR003501; PTS_IIB.
CC Pfam: PF02378; PTS_EIIC; 1.
CC DR Pfam: PF02378; PTS_IIB; 1.
CC KW Phosphotransferase system; Sugar transport; Transferase;
CC Phosphorylation; Transmembrane.
CC FT DOMAIN 1 572 EIIIC DOMAIN.
CC FT DOMAIN 1 572 EIIIB DOMAIN.
CC SEQUENCE 572 AA; 62679 MW; D943DC54B0644F CRC64;

Query Match          100.0%; Score 33; DB 1; Length 572;
Best Local Similarity 100.0%; Pred No. 1, 1e-02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGTSG 6
DB 478 GGGTSG 483

RESULT 8
PTLB_LACCA          STANDARD:      PRT:      577 AA.
ID PTLB_LACCA
AC P24400:
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE PTS SYSTEM, LACTOSE-SPECIFIC IIBC COMPONENT (EIIIC-LAC) (LACTOSE-

```

```

DE PERASE IIBC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, BC COMPONENT)
DE (EC 2.7.1.69) (EII-LAC).
GN LACE.
OS Lactobacillus casei.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxId=1582;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91093108; PubMed=2125053;
RA Alpert C.-A., Chassy B.M.;
RT "Molecular cloning and DNA sequence of lacE, the gene encoding the
RT lactose-specific enzyme II of the phosphotransferase system of
RT Lactobacillus casei. Evidence that a cysteine residue is essential
RT for sugar phosphorylation."
RT J. Biol. Chem. 265:22561-22568(1990).
CC -1- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
CC AND THE TRANSMEMBRANE CHANNEL. THE IIA DOMAIN CONTAINS THE PRIMARY
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC THE SUGAR.
CC -1- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHISTIDINE + SUGAR -
CC PROTEIN HISTIDINE + SUGAR PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: CONTAINS A PTS EIIIB DOMAIN.
CC -1- SIMILARITY: CONTAINS A PTS EIIIC DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M60851; AAA72984.1; -.
CC PIR: B23697; B23697.
CC InterPro: IPR003352; PTS_EIIC.
CC InterPro: IPR003501; PTS_IIB.
CC Pfam: PF02378; PTS_EIIC; 1.
CC DR Pfam: PF02378; PTS_IIB; 1.
CC KW Phosphotransferase system; Sugar transport; Transferase;
CC Phosphorylation; Transmembrane.
CC FT DOMAIN 1 577 EIIIC DOMAIN.
CC FT DOMAIN 1 577 EIIIB DOMAIN.
CC FT TRANSMEM 27 47 POTENTIAL.
CC FT TRANSMEM 63 83 POTENTIAL.
CC FT TRANSMEM 100 120 POTENTIAL.
CC FT TRANSMEM 133 153 POTENTIAL.
CC FT TRANSMEM 176 196 POTENTIAL.
CC FT TRANSMEM 219 239 POTENTIAL.
CC FT TRANSMEM 280 300 POTENTIAL.
CC FT TRANSMEM 326 346 POTENTIAL.
CC FT TRANSMEM 363 383 POTENTIAL.
CC FT TRANSMEM 385 406 POTENTIAL.
CC FT TRANSMEM 428 448 POTENTIAL.
CC FT TRANSMEM 476 496 POTENTIAL.
CC SEQUENCE 577 AA; 62391 MW; BE56852CAB0D76E4 CRC64;

Query Match          100.0%; Score 33; DB 1; Length 577;
Best Local Similarity 100.0%; Pred No. 1, 1e-02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGTSG 6
DB 485 GGGTSG 490

RESULT 9
GCKR_HUMAN

```

```

ID  GCR_HUMAN  STANDARD:  PRT:  625 AA.
AC  014397; Q99522;
DT  01-NOV-1997 (Rel. 35, Created)
DT  30-MAY-2000 (Rel. 39, Last sequence update)
DT  20-AUG-2001 (Rel. 40, Last annotation update)
DE  GLUCOKINASE REGULATORY PROTEIN (GLUCOKINASE REGULATOR).
GN  GCR.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX  NCBI_TaxID=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE=Hepatoblastoma;
RX  MEDLINE=96014291; PubMed=8589523;
RA  Warner J.P., Leek J.P., Intody S., Markham A.F., Bonthron D.T.;
RT  "Human glucokinase regulatory protein (GCR): cDNA and genomic
RT  cloning, complete primary structure, and chromosomal localization."
RL  Mamm. Genome 6:532-536(1995).
RN  [2]
RP  PARTIAL SEQUENCE FROM N.A., REVISIONS, AND VARIANT LEU-446.
RX  MEDLINE=98234554; PubMed=9570959;
RA  Hayward B.E., Dunlop N., Intody S., Leek J.P., Markham A.F.,
RA  Warner J.P., Bonthron D.T.;
RT  "Organization of the human glucokinase regulator gene GCR."
RL  Genomics 49:137-142(1998).
CC  -1- FUNCTION: INHIBITS GLUCOKINASE BY FORMING AN INACTIVE COMPLEX WITH
CC  THIS ENZYME.
CC  -1- TISSUE SPECIFICITY: FOUND IN LIVER AND PANCREAS. NOT DETECTED IN
CC  MUSCLE, BRAIN, HEART, THYMUS, INTESTINE, UTERUS, ADIPOSE TISSUE,
CC  KIDNEY, ADRENAL, LUNG OR SPLEEN.
CC  -1- SIMILARITY: BELONGS TO THE SIS FAMILY. GCR SUBFAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; Z48475; CAAB8367.1; -
DR  EMBL; Y09593; CAA70779.2; -
DR  EMBL; Y09592; CAB61828.1; -
DR  MIM; 600842; -
DR  InterPro: IPR001741; GCR.
DR  InterPro: IPR001347; SIS.
DR  Pfam; PF01380; SIS; 1.
DR  ProDom; PD025295; GCR; 1.
DR  PROSITE; PS01272; GCR; 1.
KM  Polymorphism.
FT  VARIANT 446 446 P->L.
FT  SEQUENCE 625 AA; 68685 MW; DE750462AC603C80 CRC64;

```

```

Query Match 100.0%; Score 33; DB 1; Length 625;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GCGTSG 6
DB 106 GCGTSG 111

```

```

RESULT 10
GCR_RAT  STANDARD:  PRT:  626 AA.
AC  007071;
DT  01-OCT-1994 (Rel. 30, Created)
DT  01-FEB-1995 (Rel. 31, Last sequence update)
DT  30-MAY-2000 (Rel. 39, Last annotation update)
DE  GLUCOKINASE REGULATORY PROTEIN (GLUCOKINASE REGULATOR).

```

```

GN  GCR.
OS  Rattus norvegicus (Rat).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX  NCBI_TaxID=10116;
RN  [1]
RP  SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC  TISSUE=Liver;
RX  MEDLINE=93236935; PubMed=7682971;
RA  Decheux M., Vandekerckhove J., van Schaftingen E.;
RT  "Cloning and sequencing of rat liver cDNAs encoding the regulatory
RT  protein of glucokinase."
RL  FEBS Lett. 321:111-115(1993).
RN  [2]
RP  REVISIONS.
RX  MEDLINE=94156054; PubMed=8112473;
RA  Decheux M., Vandekerckhove J., van Schaftingen E.;
RT  "Cloning and sequencing of rat liver cDNAs encoding the regulatory
RT  protein of glucokinase."
RL  FEBS Lett. 339:312-312(1994).
CC  -1- FUNCTION: INHIBITS GLUCOKINASE BY FORMING AN INACTIVE COMPLEX WITH
CC  THIS ENZYME.
CC  -1- TISSUE SPECIFICITY: FOUND IN LIVER. NOT DETECTED IN MUSCLE, BRAIN,
CC  HEART, TESTIS, INTESTINE OR SPLEEN.
CC  -1- SIMILARITY: BELONGS TO THE SIS FAMILY. GCR SUBFAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; X68497; CAA48511.1; -
DR  InterPro: IPR001741; GCR.
DR  InterPro: IPR001347; SIS.
DR  Pfam; PF01380; SIS; 1.
DR  ProDom; PD025295; GCR; 1.
DR  PROSITE; PS01272; GCR; 1.
FT  INT-MET 0
FT  SEQUENCE 626 AA; 68786 MW; 60A6A57EBACD1FE3 CRC64;

```

```

Query Match 100.0%; Score 33; DB 1; Length 626;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GCGTSG 6
DB 105 GCGTSG 110

```

```

RESULT 11
K2C3_HUMAN  STANDARD:  PRT:  629 AA.
AC  P12035;
DT  01-OCT-1989 (Rel. 12, Created)
DT  01-OCT-1989 (Rel. 12, Last sequence update)
DT  20-AUG-2001 (Rel. 40, Last annotation update)
DE  KERATIN, TYPE II CYTOSKELETAL 3 (CYTOKERATIN 3) (K3) (CK3) (65 KDA
DE  CYTOKERATIN).
GN  KRT3.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX  NCBI_TaxID=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=87254239; PubMed=2439698;
RA  Klinge E.M., Sylvestre Y.R., Friedberg I.M., Blumenberg M.;
RT  "Evolution of keratin genes: different protein domains evolve by
RT  different pathways."

```

RL J. Mol Evol. 24:319-329(1987).  
 RN [2]  
 RP VARIANT MCD LYS-509.  
 RA MEDLINE-97315826; PubMed-9171831;  
 RA Irvine A.D., Gorden L.D., Swenson O., Swenson B., Moore J.E.,  
 RA Frazer D.G., Smith F.J.D., Knowlton R.G., Christophers E.,  
 RA Rochelle R., Uitto J., McLean W.H.I.;  
 RA "Mutations in cornea-specific keratin K3 or K12 genes cause  
 RT Meesmann's corneal dystrophy.";  
 RL Nat. Genet. 16:184-187(1997).  
 CC -1- SODUNIT: HETEROTRIMER OF TWO TYPE I AND TWO TYPE II KERATINS.  
 CC KERATIN 3 ASSOCIATES WITH KERATIN 12.  
 CC -1- TISSUE SPECIFICITY: CORNEA-SPECIFIC.  
 CC -1- DISEASE: DEFECTS IN KRT3 ARE A CAUSE OF MEESMANN CORNEAL DYSTROPHY  
 CC (MCD), AN AUTOSOMAL DOMINANT DISEASE THAT CAUSES FRAGILITY OF THE  
 CC ANTERIOR CORNEAL EPITHELIUM. PATIENTS ARE USUALLY ASYMPTOMATIC  
 CC UNTIL ADULTHOOD WHEN RUPTURE OF THE CORNEAL MICROCYSTS MAY CAUSE  
 CC EROSIONS, PRODUCING CLINICAL SYMPTOMS SUCH AS PHOTOPHOBIA, CONTRACT  
 CC LENS INTOLERANCE AND INTERMITTENT DIMINUTION OF VISUAL ACUTY.  
 CC RARELY, SUBEPITHELIAL SCARRING CAUSES IRREGULAR CORNEAL  
 CC ASTIGMATISM AND PERMANENT VISUAL IMPAIRMENT. HISTOLOGICAL  
 CC EXAMINATION SHOWS A DISORGANIZED AND THICKENED EPITHELIUM WITH  
 CC WIDESPREAD CYTOPLASMIC VACUOLIZATION AND NUMEROUS SMALL, ROUND,  
 CC DEBRIS-LADEN INTRAEPITHELIAL CYSTS.  
 CC -1- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND  
 CC MICROFIBILLAR KERATIN: I (ACIDIC; 40-55 KDA) [K9 TO K20] AND II  
 CC (NEUTRAL TO BASIC; 56-70 KDA) [K1 TO K8].  
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X05418; CAA28991.1; -.  
 DR EMBL: X05419; CAA28992.1; ALT\_SEQ.  
 DR EMBL: X03420; CAA28993.1; ALT\_SEQ.  
 DR EMBL: X03420; CAA28994.1; ALT\_SEQ.  
 DR EMBL: X05420; CAA28995.1; ALT\_SEQ.  
 DR EMBL: X05421; CAA28996.1; ALT\_SEQ.  
 DR PIR: A29666; A29666.  
 DR MIM: 148043; -.  
 DR MIM: 122100; -.  
 DR InterPro: IPR001664; IF.  
 DR InterPro: IPR003054; Keratin\_II.  
 DR Pfam: PF00038; filament.1.  
 DR PRINTS: PR01276; TYPE2KERATIN.  
 DR PROSITE: PS00226; IF, 1.  
 KW Intermediate filament; Coiled coil; Keratin; Phosphorylation;  
 KW Disease mutation.  
 FT DOMAIN 1 197 HEAD.  
 FT DOMAIN 198 509 ROD.  
 FT DOMAIN 510 629 TAIL.  
 FT DOMAIN 198 233 COIL 1A.  
 FT DOMAIN 234 254 LINER 1.  
 FT DOMAIN 255 346 COIL 1B.  
 FT DOMAIN 347 370 LINER 12.  
 FT DOMAIN 371 509 COIL 2.  
 FT MOD\_RES 56 56 PHOSPHORYLATION (BY SIMILARITY).  
 FT VARIANT 509 509 E->K (IN MCD).  
 SEQUENCE 629 AA; 64511 MW; 2E746B19A828BCD9 CRC64;  
 /FTID=VAR\_003868.

```

Query Match      100.0%;   Score 33;   DB 1;   Length 629;
Best Local Similarity 100.0%;   Pred. No. 1.2e+02;
Matches 6;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;
QY      1 GGGTSG 6
        |||||

```

```

Db      578 GGTSG 583

RESULT 12
UL52_HCMVA
ID      UL52_HCMVA      STANDARD:      PRT:      668 AA.
AC      P16793;
DT      01-AUG-1990 (Rel. 15, Created)
DT      01-AUG-1990 (Rel. 15, Last sequence update)
DT      01-OCT-1996 (Rel. 34, Last annotation update)
DE      PROBABLE MAJOR ENVELOPE GLYCOPROTEIN UL52 (HFRL1 PROTEIN),
GN      UL52.
OS      Human cytomegalovirus (strain AD169).
OC      Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC      Betaherpesvirinae; Cytomegalovirus.
OX      NCBI_TaxID=10360;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=90269039; PubMed=2161319;
RA      Chee M.S., Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R.,
RA      Horsnell T., Hutchinson C.A. III, Kouzarides T., Martignetti J.A.,
RA      Priddele E., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.;
RT      "Analysis of the protein-coding content of the sequence of human
RT      cytomegalovirus strain AD169."
RL      Curr. Top. Microbiol. Immunol. 154:125-169(1990).
RL      -1. SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL32,
RL      EBV-1 28, EBV BFLF1, HCMV UL52, AND VZV 26.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      at the European Bioinformatics Institute. There are no restrictions on
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/isb-sib.ch).
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: X17403; CAA5411.1; -.
DR      PIR: S09815; Q08EM1.
DR      InterPro: IPR002597; Herpes_env.
DR      Pfam: PF01673; Herpes_env; 1.
DR      GlycoProtein.
FT      CARBOHYD 142      142      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 145      145      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 295      295      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 453      453      N-LINKED (GLCNAC. . .) (POTENTIAL).
SO      SEQUENCE 668 AA; 74120 MW; 29f675f40f3ed818 crc64.

```

```

Query Match          100.0%; Score 33; DB 1; Length 668;
Best Local Similarity 100.0%; Pred. NO. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 GGGTSG 6
        |||||
Db       422 GGGTSG 427

RESULT 13
SP87_DICDI
ID SP87_DICDI STANDARD; PRT; 677 AA.
AC P56643:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE SP87 COAT PROTEIN SP87 PRECURSOR (PL3 PROTEIN).
GN PSPD.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
NCBI_OX [1]_taxid=44689;
RN
RP
RC SEQUENCE FROM N.A.
RC STRAIN=AX3;
RX MEDLINE=94229356; PubMed=8174787;

```

RA Yoder B.K., Mao J., Erdos G.W., West C.M., Blumberg D.D.:  
 RT "Identification of a new spore coat protein gene in the cellular  
 RL slime mold Dictyostelium discoideum.";  
 CC Dev. Biol. 163:49-65(1994).  
 CC -1- FUNCTION: MAY CONTRIBUTE TO THE STRUCTURE OF THE COAT AT THE  
 CC INTERFACE BETWEEN THE MIDDLE, CELLOUSIC LAYER AND THE OUTER,  
 CC ELECTRON-DENSE, PROTEINACOUS LAYER.  
 CC -1- SUBCELLULAR LOCATION: ACCUMULATES SPECIFICALLY IN REGULATED  
 CC SECRETORY VESICLES OF PRESPORE CELLS (PRESPORE VESICLES). THE  
 CC PROTEIN LATER ACCUMULATES EXTRACELLULARLY IN THE SPORE COAT.  
 CC -1- INDUCTION: BY C-AMP.  
 CC -1- PTM: DISULFIDE BONDING IS IMPORTANT FOR ASSOCIATING SPB7 WITH THE  
 CC COAT.  
 CC -1- PTM: PHOSPHORYLATED ON SERINE RESIDUES.  
 CC -1- SIMILARITY: CONTAINS 12 PRESPORE MOTIFS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U25144; AAA73515.1; -;  
 DR HSSP: P06620; 11NA.  
 DR DClcydb: DD02054; PSPD.  
 DR InterPro: IPR003645; FOLN.  
 DR SMART: SM00274; FOLN; 5.  
 KW Glycoprotein; Repeat; Sporulation; Signal.  
 FT SIGNAL 1 21  
 FT CHAIN 1 21  
 FT REPEAT 189 201 SPORE COAT PROTEIN SPB7.  
 FT REPEAT 232 244 PRESPORE MOTIF 1.  
 FT REPEAT 276 288 PRESPORE MOTIF 2.  
 FT REPEAT 316 328 PRESPORE MOTIF 3.  
 FT REPEAT 356 368 PRESPORE MOTIF 4.  
 FT REPEAT 390 402 PRESPORE MOTIF 5.  
 FT REPEAT 418 430 PRESPORE MOTIF 6.  
 FT REPEAT 450 462 PRESPORE MOTIF 7.  
 FT REPEAT 513 525 PRESPORE MOTIF 8.  
 FT REPEAT 543 555 PRESPORE MOTIF 9.  
 FT REPEAT 571 583 PRESPORE MOTIF 10.  
 FT REPEAT 601 613 PRESPORE MOTIF 11.  
 FT REPEAT 648 660 PRESPORE MOTIF 12.  
 FT DOMAIN 148 156 5 X 9 AA REPEATS OF G-G-S-S-S-G-T-S.  
 FT REPEAT 157 165 1-1 (APPROXIMATE).  
 FT REPEAT 166 174 1-2.  
 FT REPEAT 175 183 1-3.  
 FT REPEAT 184 192 1-4 (APPROXIMATE).  
 FT DOMAIN 268 387 1-5 (APPROXIMATE).  
 FT REPEAT 308 347 3 X 40 AA APPROXIMATE REPEATS.  
 FT REPEAT 348 387 2-1.  
 FT REPEAT 391 628 2-2.  
 FT DOMAIN 391 417 8 X 26 AA APPROXIMATE REPEATS.  
 FT REPEAT 419 445 3-1.  
 FT REPEAT 451 477 3-2.  
 FT REPEAT 480 506 3-3.  
 FT REPEAT 514 540 3-4.  
 FT REPEAT 544 570 3-5.  
 FT REPEAT 572 598 3-6.  
 FT REPEAT 602 628 3-7.  
 FT REPEAT 628 648 3-8.  
 FT DOMAIN 141 185 GLY/SER/THR-RICH.  
 FT DOMAIN 256 271 GLY/SER/THR-RICH.  
 FT DOMAIN 342 351 GLY/SER/THR-RICH.  
 FT DOMAIN 174 183 POLY-SER.  
 FT DOMAIN 445 448 POLY-PRO.  
 FT CAROHD 72 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT SEQUENCE 677 AA; 71804 MW; EF6E29CF57E78D5 CRC64;

Query Match

100.0%; Score 33; DB 1; Length 677;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
 Db 256 GCGTSG 261

## RESULT 14

NH48\_CAEEL STANDARD; PRT; 780 AA.  
 ID NH48\_CAEEL  
 AC Q94407;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DE NUCLEAR HORMONE RECEPTOR FAMILY MEMBER NHR-48.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_Taxid=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA White S.;  
 RL Submitted (Aug-1996) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: ORPHAN NUCLEAR RECEPTOR.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: Z79604; CAB01900.1; -;  
 DR HSSP: P10826; 1HRA.  
 DR WormPep: ZK662.3; CE18456.  
 DR InterPro: IPR001628; Zf-C4.  
 DR Pfam: PF00105; Zf-C4; 1.  
 DR SMART: SM00399; Znf\_C4; 1.  
 DR PROSITE: PS00031; NUCLEAR\_RECEPTOR; 1.  
 KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;

KW DNA\_BIND 100 165 NUCLEAR RECEPTOR-TYPE.  
 FT ZN\_FING 100 120 C4-TYPE.  
 FT ZN\_FING 136 160 C4-TYPE.  
 SQ SEQUENCE 780 AA; 87373 MW; 74922FB93F7D8B40 CRC64;

Query Match 100.0%; Score 33; DB 1; Length 780;

Best Local Similarity 100.0%; Pred. No. 1.5e+02; Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
 Db 64 GCGTSG 69

## RESULT 15

KF3C\_RAT STANDARD; PRT; 796 AA.  
 ID KF3C\_RAT

AC O55165; O88657;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE KINESIN-LIKE PROTEIN KIF3C.  
 GN KIF3C.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBL\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98155212; PubMed=9487132;  
 RA Muretan V., Abramson T., Lyass A., Winter D., Porro E., Hong F.,  
 RA Chamberlin N.L., Schnapp B.J.;  
 RT "KIF3C and KIF3A form a novel neuronal heteromeric kinesin that  
 RT associates with membrane vesicles.";  
 RL Mol. Biol. Cell 9:637-652(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99023472; PubMed=9808286;  
 RA Faure K., Gruber D., Bulinski J.C.;  
 RT "Identification of kinesin-like molecules in myogenic cells.";  
 RL Eur. J. Cell Biol. 77:27-34(1998).  
 CC -1- FUNCTION: MICROTUBULE-BASED ANTEROGRADE TRANSLOCATOR FOR  
 CC MEMBRANOUS ORGANELLES.  
 CC -1- SUBUNIT: HETERODIMER OF KIF3A AND KIF3C.  
 CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN  
 CC II SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AJ223599; CA11465.1; -;  
 DR EMBL: AF083330; AAC33291.1; -;  
 DR InterPro: IPR001752; Kinesin.  
 DR Pfam: PF00225; Kinesin; 2.  
 DR PRINTS: PR00380; KINESINHEAVY.  
 DR SMART: SM00129; KISC.1.  
 DR PROSITE: PS00411; KINESIN\_MOTOR\_DOMAIN1; 1.  
 DR Motor protein; Microtubules; ATP-binding; Colled coil; Neurone.  
 FT DOMAIN 1 377  
 FT DOMAIN 378 632  
 FT DOMAIN 633 793  
 FT NP\_BIND 97 104  
 FT DOMAIN 271 278  
 FT DOMAIN 440 443  
 FT DOMAIN 772 775  
 FT CONFLICT 352 352  
 FT CONFLICT 562 563  
 SO SEQUENCE 796 AA; 89815 MW; 935A79A01F8BDC0D CRC64;  
 ML -> DD (IN REF. 2).  
 ML -> IV (IN REF. 2).  
 Query Match 100.0%; Score 33; DB 1; Length 796;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 GGCTSG 6  
 DB 276 GGCTSG 281  
 RESULT 16  
 CLAB\_LYCES  
 ID CLAB\_LYCES STANDARD: PRT; 923 AA.  
 AC P31542;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE ATP-DEPENDENT CLP PROTEASE ATP-BINDING SUBUNIT CLPA HOMOLOG CD4B  
 DE PRECURSOR.  
 GN CD4B.  
 OS Lycopersicon esculentum (Tomato).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

OX NCBL\_TaxID=4081;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90239044; PubMed=2185473;  
 RA Gottesman S., Squires C., Pichersky E., Carrington M., Hobbs M.,  
 RA Matlick J.S., Dalrymple B., Kurumitsu H., Shiroza T., Foster T.,  
 RA Clark W.P., Ross B., Squires C.L., Maurizi M.R.;  
 RT "Conservation of the regulatory subunit for the Clp ATP-dependent  
 RT protease in prokaryotes and eukaryotes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:3513-3517(1990).  
 CC -1- FUNCTION: MAY INTERACT WITH A CLP-LIKE PROTEASE INVOLVED IN  
 CC DEGRADATION OF DENATURED PROTEINS IN THE CHLOROPLAST.  
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.  
 CC -1- SIMILARITY: BELONGS TO THE CLPA/CLPB FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M32604; AAA34161.1; -;  
 DR PIR: B35905; B35905.  
 DR InterPro: IPR003593; AAA.  
 DR InterPro: IPR001939; AAA\_subfam.  
 DR InterPro: IPR001270; CLP\_AB.  
 DR InterPro: IPR001943; UVR.  
 DR Pfam: PF00004; AAA; 2.  
 DR Pfam: PF02151; UVR; 1.  
 DR PRINTS: PR00300; CLPPROTEASEA.  
 DR SMART: SM00382; AAA; 2.  
 DR PROSITE: PS00870; CLPAB\_1; 1.  
 DR PROSITE: PS00871; CLPAB\_2; 1.  
 DR Chapterone; ATP-binding; Repeat; Chloroplast; Transit peptide.  
 FT FT 1 923  
 FT CHAIN ? 923  
 FT FT 1 923  
 FT FT 255 502  
 FT DOMAIN 569 760  
 FT NP\_BIND 300 307  
 FT NP\_BIND 643 650  
 FT NP\_BIND 643 650  
 FT ATP (POTENTIAL).  
 FT ATP (POTENTIAL).  
 SO SEQUENCE 923 AA; 102241 MW; 26F9DCFE6974E65 CRC64;  
 Query Match 100.0%; Score 33; DB 1; Length 923;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 GGCTSG 6  
 DB 244 GGCTSG 249  
 RESULT 17  
 CLAA\_LYCES  
 ID CLAA\_LYCES STANDARD: PRT; 926 AA.  
 AC P31541;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE ATP-DEPENDENT CLP PROTEASE ATP-BINDING SUBUNIT CLPA HOMOLOG CD4A  
 DE PRECURSOR.  
 GN CD4A.  
 OS Lycopersicon esculentum (Tomato).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
 OX NCBL\_TaxID=4081;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90239044; PubMed=2185473;

RA Gottesman S., Squires C., Pichersky E., Carrington M., Hobbs M.,  
 RA Matlick J.S., Dalrymple B., Kurimatsu H., Shiroza T., Foster T.,  
 RA Clark W.P., Rose B., Squires C.L., Maurizi M.R.;  
 RT "Conservation of the regulatory subunit for the Clp ATP-dependent  
 RT protease in prokaryotes and eukaryotes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:3513-3517(1990).  
 CC -1- FUNCTION: MAY INTERACT WITH A CLP-P-LIKE PROTEASE INVOLVED IN  
 CC DEGRADATION OF DENATURED PROTEINS IN THE CHLOROPLAST.  
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.  
 CC -1- SIMILARITY: BELONGS TO THE CLPA/CLPB FAMILY. CLPC SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M32603; AAA34160.1; -.  
 DR PIR: A35905; A35905.  
 DR InterPro: IPR003593; AAA.  
 DR InterPro: IPR001939; AAA\_subfam.  
 DR InterPro: IPR001270; CLP\_AB.  
 DR InterPro: IPR001943; UVR.  
 DR Pfam: PF00004; AAA; 2.  
 DR Pfam: PF02151; UVR; 1.  
 DR PRINTS: PR00300; CLPPTROTEASEA.  
 DR SMART: SM00382; AAA; 2.  
 DR PROSITE: PS00870; CLPAB\_1; 1.  
 DR PROSITE: PS00871; CLPAB\_2; 1.  
 DR Chaperone: ATP-binding; Repeat: Chloroplast; Transit peptide.  
 KW Chaperone: ATP-binding; Repeat: Chloroplast; Transit peptide.  
 FT TRANSIT 1 ? 926  
 FT CHAIN 1 ? 926  
 FT DOMAIN 110 216  
 FT REPEAT 110 141  
 FT REPEAT 185 216  
 FT DOMAIN 257 504  
 FT DOMAIN 572 763  
 FT NP\_BIND 302 309  
 FT NP\_BIND 646 653  
 FT NP\_BIND 102556 MW; 30FA31F892E0916C CRC64;  
 SQ SEQUENCE 926 AA; 102556 MW; 30FA31F892E0916C CRC64;  
 Query Match 100.0%; Score 33; DB 1; Length 926;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GCGTSG 6  
 DB 246 GCGTSG 251  
 RESULT 18  
 YL55\_CAEEL  
 ID YL55\_CAEEL STANDARD; PRT; 32 AA.  
 AC P34435;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE HYPOTHEICAL 3.4 KDA PROTEIN F44E2.5 IN CHROMOSOME III.  
 GN F44E2.5.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Pelodermidae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RX NCBI\_TaxID=6239;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berts M.,  
 RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,

RA Craxton M., Dear S., Du Z., Durbin R., Favella A., Fraser A.,  
 RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,  
 RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,  
 RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,  
 RA Stims M., Smaildon N., Smith A., Smith M., Sonhammer E., Staden K.,  
 RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,  
 RA Watson R., Watson A., Weinstock L., Wilkinson-Spratt J.,  
 RA Wooldman P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans.";  
 RL Nature 368:32-38(1994).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: L23646; AK67233.1; -.  
 DR Wormpep: F44E2.5; CE00183.  
 DR Hypothetical protein.  
 SQ SEQUENCE 32 AA; 3393 MW; 0744BB516215EC75 CRC64;  
 Query Match 90.9%; Score 30; DB 1; Length 32;  
 Best Local Similarity 83.3%; Pred. No. 21;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GCGTSG 6  
 DB 11 GCGTAG 16  
 RESULT 19  
 NRDI\_ECOLI  
 ID NRDI\_ECOLI STANDARD; PRT; 136 AA.  
 AC 047415; P77025;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE NRDI PROTEIN.  
 GN NRDI OR B2674 OR Z3976 OR ECS3537.  
 OS Escherichia coli, and  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562, 83334;  
 RX NCBI\_TaxID=562, 83334;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RX MEDLINE=96417857; PubMed=8820648;  
 RA Jordan A., Aragaal E., Gilbert I., Barbe J.;  
 RT "Promoter identification and expression analysis of Salmonella  
 RT typhimurium and Escherichia coli uidP operons encoding one of two  
 RT class I ribonucleotide reductases present in both bacteria.";  
 RL Mol. Microbiol. 19:777-790(1996).  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RX MEDLINE=97349980; PubMed=9205837;  
 RA

RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,  
 RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,  
 RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,  
 RA Oshima T., Oyama S., Saico N., Sampei G., Satoh Y., Sivasubram S.,  
 RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,  
 RA Yamagata S., Horinouchi T.,  
 RT Construction of a contiguous 874-kb sequence of the Escherichia coli  
 RT -R12 genome corresponding to 50,0-68.8 mln on the linkage map and  
 RT analysis of its sequence features.";  
 RL DNA Res. 4:91-113(1997).  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-0157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE-21074935; PubMed-11206551;  
 RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic Escherichia coli 0157:H7.";  
 RL Nature 409:529-533(2001).  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-0157:H7 / RIMD 0509952;  
 RX MEDLINE-21156231; PubMed-11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Ikeda T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
 RA Kihara S., Shiba T., Hattori M., Shingawa H.;  
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
 RT 0157:H7 and genomic comparison with a laboratory strain K-12.";  
 RL DNA Res. 8:11-22(2001).  
 CC  
 CC -1- FUNCTION: NOT KNOWN; PROBABLY INVOLVED IN RIBONUCLEOTIDE REDUCTASE  
 CC FUNCTION.  
 CC  
 CC -1- SIMILARITY: BELONGS TO THE NROI FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X79787; CAA56185.1; -;  
 DR EMBL; AE000352; AAC75721.1; -;  
 DR EMBL; D90891; BAA16538.1; ALT\_INIT.  
 DR EMBL; AE005496; AAG57784.1; -;  
 DR EMBL; AP002562; BAB36960.1; -;  
 DR Ecogene; EG13287; nrcl.  
 DR InterPro; IPR003707; Nrcl.  
 DR Pfam; PF02555; Nrcl; 1.  
 DR Complete proteome.  
 KW  
 SO SEQUENCE 136 AA; 15340 MW; 2614D080D18EBCBB CRC64;

Query Match 90.9%; Score 30; DB 1; Length 136;  
 Best Local Similarity 83.3%; Pred. No. 87;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGGTSG 6  
 DB 51 GGGTAG 56

RESULT 20  
 YJF5\_YEAST STANDARD: PRT; 245 AA.  
 AC P47044;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE HYPOTHETICAL 26.9 KDA PROTEIN IN BTN1-PEP8 INTERGENIC REGION.  
 GN YJ055W OR J1148.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Pohl T.M., Aljinovic G.;  
 RL Submitted (SEP-1995) to the EMBL/Genbank/DBJ databases.  
 CC  
 CC -1- SIMILARITY: TO R.FASCIANS HYPOTHETICAL 21.1 KDA PROTEIN IN  
 CC FASCINATION LOCUS (AC P46378).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; Z49330; CAA89346.1; -;  
 DR SGD; S0003591; YJ055W.  
 DR Hypothetical protein; ATP-binding.  
 KW NP\_BIND  
 FT BIND 13  
 SO SEQUENCE 245 AA; 26864 MW; 0C49E7FC4F1E7CA2 CRC64;

Query Match 90.9%; Score 30; DB 1; Length 245;  
 Best Local Similarity 83.3%; Pred. No. 1.5e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGGTSG 6  
 DB 56 GGGTTC 61

RESULT 21  
 IF31\_HUMAN STANDARD: PRT; 258 AA.  
 AC 073822;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE EUKARYOTIC TRANSLATION INITIATION FACTOR 3 SUBUNIT 1 (EIF-3 ALPHA)  
 DE (EIF3 P35).  
 GN EIF3S1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE-99041954; PubMed-9822659;  
 RA Block K.L., Vornlocher H.-P., Hershey J.W.B.;  
 RT "Characterization of cDNAs encoding the p44 and p35 subunits of human  
 RT translation initiation factor eif3.";  
 RL J. Biol. Chem. 273:31901-31906(1998).  
 CC  
 CC -1- FUNCTION: BINDS TO THE 40S RIBOSOME AND PROMOTES THE BINDING OF  
 CC METHIONYL-TRNAI AND MRNA.  
 CC  
 CC -1- SUBUNIT: EIF-3 IS COMPOSED OF AT LEAST 10 DIFFERENT SUBUNITS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U97670; AAC76729.1; -;  
 DR MIM; 603910; -;  
 KW Initiation factor; Protein biosynthesis.

FT DOMAIN 2 8 POLY-ALA.  
FT DOMAIN 29 32 POLY-GLY.  
FT DOMAIN 53 57 POLY-ASP.  
FT DOMAIN 218 224 POLY-LYS.  
SQ SEQUENCE 258 AA; 28990 MW; F01E75741473058B CRC64;

Query Match 90.9%; Score 30; DB 1; Length 258;  
Best Local Similarity 83.3%; Pred. No. 1.6e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
111111  
Db 30 GCGTAG 35

RESULT 22  
EUTJ\_ECOLI STANDARD; PRT; 278 AA.  
ID EUTJ\_ECOLI P72277;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE ETHANOLAMINE UTILIZATION PROTEIN EUTJ.  
GN EUTJ OR B2454.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.

OC NCB1\_TaxID=562;  
OX NCB1\_TaxID=562;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;

RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Valdes J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shaoy Y.;  
RT "The complete genome sequence of Escherichia coli K-12";  
RL Science 277:1453-1474(1997).  
RN (2)

RP SEQUENCE FROM N.A.

RC STRAIN-K12;  
RX MEDLINE=97349980; PubMed=9205837;  
RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,  
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,  
RA Mitsuhashi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,  
RA Oshima T., Oyama S., Saito N., Sempel G., Satoh Y., Sivasubram S.,  
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,  
RA Yamagata S., Horiuchi T.;  
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli  
RT K-12 genome corresponding to 50.0-68.8 min on the linkage map and  
RT DNA Res. 4:91-113(1997).  
RN (1)

RL DNA Res. 4:91-113(1997).  
RN (2)  
RP SEQUENCE FROM N.A.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC

CC EMBL: AE000332; AAC75507.1; -  
DR EMBL: D90874; BAA16332.1; -  
DR Ecocore: EC14184; eutJ.  
KW Complete proteome.  
SQ SEQUENCE 278 AA; 30062 MW; 8F92B5DE54ED9FF CRC64;

Query Match 90.9%; Score 30; DB 1; Length 278;  
Best Local Similarity 83.3%; Pred. No. 1.7e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
111111  
Db 145 GCGTTG 150

RESULT 23  
EUTJ\_SALTY STANDARD; PRT; 279 AA.  
ID EUTJ\_SALTY P41794;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE ETHANOLAMINE UTILIZATION PROTEIN EUTJ.  
GN EUTJ.

OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OX NCB1\_TaxID=602;  
RN (1)

RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC 14028S;  
RX MEDLINE=95173114; PubMed=7868611;  
RA Stojiljkovic I., Baemler A.J., Heffron F.;  
RT "The 17-gene ethanolamine (eut) operon of Salmonella typhimurium  
RT encodes five homologues of carboxysome shell proteins";  
RL J. Bacteriol. 181:5317-5329(1999).  
RN (2)  
RP SEQUENCE FROM N.A.

RC STRAIN-LT2;  
RX MEDLINE=99395039; PubMed=10464203;  
RA Kofoid E.C., Rappleye C.A., Stojiljkovic I., Roth J.R.;  
RT "The 17-gene ethanolamine (eut) operon of Salmonella typhimurium  
RT encodes five homologues of carboxysome shell proteins";  
RL J. Bacteriol. 181:5317-5329(1999).  
RN (1)

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC

DR EMBL: U18560; AAA80210.1; -  
DR EMBL: AF093749; AAC78119.1; -  
DR StGene: SG10554; eutJ  
DR SEQUENCE 279 AA; 30018 MW; 28B8CD89141D8090 CRC64;

Query Match 90.9%; Score 30; DB 1; Length 279;  
Best Local Similarity 83.3%; Pred. No. 1.8e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
111111  
Db 146 GCGTTG 151

RESULT 24  
OTX3\_BRARE STANDARD; PRT; 338 AA.  
ID OTX3\_BRARE O90267;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE HOMEBOX PROTEIN OTX3 (ZOTX3).  
GN OTX3.

OS Brachydanio rerio (zebrafish) (Zebra danio).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;

CC Cypriniformes; Cyprinidae; Rasbora; Danio.  
 OX NCBI\_TaxID=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-OREGON AB;  
 RX MEDLINE=95206106; PubMed=7898305;  
 RA Mori H., Miyazaki Y., Morita T., Nitta H., Mishina M.;  
 RT "Different spatio-temporal expressions of three otx homeoprotein  
 transcripts during zebrafish embryogenesis.";  
 RL Brain Res. Mol. Brain Res. 27:221-231(1994).  
 CC -1- FUNCTION: MAY PLAY A ROLE IN VERY EARLY EMBRYOGENESIS,  
 CC GASTRULATION, AND THE DEVELOPMENT AND SUBDIVISION OF THE  
 CC DIENCEPHALON AND THE MIDBRAIN. MAY PLAY A ROLE IN THE ORGANIZER  
 CC FUNCTION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).  
 CC DEVELOPMENTAL STAGE: DETECTED ALONG THE MARGINAL ZONE OF SYMMETRIC  
 CC EMBRYOS AT 5 HRS OF DEVELOPMENT AND FOUND AT THE SHIELD, A PRIMARY  
 CC MORPHOLOGICAL ASYMMETRY AT 6 HRS OF DEVELOPMENT. FOUND IN TWO  
 CC STRIPES AT THE POSTERIOR SIDE OF THE EYE RUDIMENTS AT 12 HRS OF  
 CC DEVELOPMENT, DISTRIBUTED IN THE DIENCEPHALON, MIDBRAIN AND THE  
 CC EPIPHYSIS AT 18 HRS OF DEVELOPMENT, AND FOUND IN THE DIENCEPHALON  
 CC AND THE MIDBRAIN AT 24 HRS OF DEVELOPMENT.  
 CC -1- SIMILARITY: BELONGS TO THE PAIRED FAMILY OF HOMEOBOX PROTEINS.  
 CC "BLOID" SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: D26174; BAA05160.1; -  
 DR HSSP: P06601; 1FJL.  
 DR ZFIN: ZDB-GENE-980526-27; ctx3.  
 DR InterPro: IPR001356; Homeobox.  
 DR InterPro: IPR003025; Ctx-TF.  
 DR Pfam: PF00046; homeobox; 1.  
 DR PRINTS: PRO0024; HOMEBOX.  
 DR PRINTS: PRO1255; OTXHOMEBOX.  
 DR SMART: SM00389; HOX; 1.  
 DR PROSITE: PS00027; HOMEBOX\_1; 1.  
 DR PROSITE: PS0071; HOMEBOX\_2; 1.  
 DR Homeobox; DNA-binding; Developmental protein; Nuclear protein.  
 FT DNA\_BIND 38 97 HOMEBOX.  
 FT DOMAIN 95 99 POLY-GLN.  
 FT DOMAIN 134 144 POLY-SER.  
 FT DOMAIN 158 162 POLY-SER.  
 FT SEQUENCE 338 AA; 36145 MW; ACPA6C2A43A69014 CRC64;  
 QY 1 GCGTSG 6  
 DB 203 GCGTTG 208  
 Query Match 90.9%; Score 30; DB 1; Length 338;  
 Best Local Similarity 83.3%; Pred. No. 2.1e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OS Streptomyces collinus.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=42684;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-DSM 40733;  
 RA Mikulik K., Zhulanova E.;  
 RT "Characterization of dcw cluster from Streptomyces.";  
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: CELL WALL FORMATION. CATALYZES THE TRANSFER OF A GLCNAC  
 CC SUBUNIT ON UNDECAPRENYL-PYROPHOSPHORYL-MURNAIC-PENTAPEPTIDE (LIPID  
 CC INTERMEDIATE I) TO FORM UNDECAPRENYL-PYROPHOSPHORYL-MURNAIC-  
 CC (PENTAPEPTIDE)GLCNAC (LIPID INTERMEDIATE II) (BY SIMILARITY).  
 CC -1- PATHWAY: LAST STEP OF PEPTIDOGLYCAN BIOSYNTHESIS.  
 CC -1- SUBCELLULAR LOCATION: MEMBRANE ASSOCIATED (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE MURG FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: AF10367; AAD26629.1; -  
 DR Transferase: Glycosyltransferase; Cell division; Cell wall;  
 KW Membrane; Peptidoglycan synthesis.  
 SQ SEQUENCE 362 AA; 38847 MW; 0BBA131F25D5122 CRC64;

Query Match 90.9%; Score 30; DB 1; Length 362;  
 Best Local Similarity 83.3%; Pred. No. 2.3e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GCGTSG 6  
 DB 7 GCGTAG 12  
 RESULT 26  
 MURG\_STRCO STANDARD; PRT; 364 AA.  
 ID MURG\_STRCO  
 AC 092BA5;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE UDP-N-ACETYLGLUCOSAMINE--N-ACETYLURAMYL-(PENTAPEPTIDE)  
 DE PYROPHOSPHORYL-UNDECAPRENYL-N-ACETYLURAMYL-(PENTAPEPTIDE)  
 DE (EC 2.4.1.-) (UNDECAPRENYL-P-MURNAIC-PENTAPEPTIDE-UDPGLCNAC GLCNAC  
 DE TRANSFERASE).  
 GN MURG OR SC4A10.17C.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RC STRAIN-A3(2);  
 RA Kuennen R.A., Stadelmaier B.T., McCormick J.R.;  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: CELL WALL FORMATION. CATALYZES THE TRANSFER OF A GLCNAC  
 CC SUBUNIT ON UNDECAPRENYL-PYROPHOSPHORYL-MURNAIC-PENTAPEPTIDE (LIPID  
 CC INTERMEDIATE I) TO FORM UNDECAPRENYL-PYROPHOSPHORYL-MURNAIC-  
 CC (PENTAPEPTIDE)GLCNAC (LIPID INTERMEDIATE II) (BY SIMILARITY).  
 CC -1- PATHWAY: LAST STEP OF PEPTIDOGLYCAN BIOSYNTHESIS.

```
CC -1- SUBCELLULAR LOCATION: MEMBRANE ASSOCIATED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE MORG FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U10879; AAD10537.1; -.
DR EMBL: AL109663; CAB51993.1; -.
KM Transferrase: Glycosyltransferase; Cell division; Cell wall;
KW Membrane; Peptidoglycan synthesis.
SQ SEQUENCE 364 AA; 38842 MW; D600C5F94E9202E2 CRC64;

Query Match          90.9%; Score 30; DB 1; Length 364;
Best Local Similarity 83.3%; Pred. No. 2.3e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCTSG 6
DB 7 CGGTAG 12

RESULT 27
MAF2_RAT          STANDARD; PRT; 369 AA.
ID MAF2_RAT
AC P54844;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE TRANSCRIPTION FACTOR MAF2 (PROTO-ONCOGENE C-MAF).
GN MAF2 OR MAF.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WISTAR;
RX MEDLINE=97190228; PubMed=9038383;
RA Kuboki Y., Imaki J., Yoshida K., Ogata A., Matsushima-Hibaya Y.,
RA Kuboki Y., Nishizawa M., Nishi S.;
RT *Rat maf related genes: specific expression in chondrocytes, lens and
RT spinal cord.
RL Oncogene 14:745-750(1997).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: BELONGS TO THE BZIP FAMILY. MAF SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U56242; AAB50063.1; -.
DR HSSP: P05412; IJUN.
DR InterPro: IPR001871; bZIP.
DR SMART: SM00338; BRLZ; 1.
KW Proto-oncogene; Transcription regulation; DNA-binding; Activator;
KW Nuclear protein.
FT DOMAIN 139 146 POLY-ALA.
FT DOMAIN 169 173 POLY-ALA.
FT DOMAIN 180 187 POLY-HIS.
FT DOMAIN 191 194 POLY-HIS.
FT DOMAIN 212 220 POLY-GLY.
FT DOMAIN 225 235 POLY-GLY.
FT DNA_BIND 285 311 BASIC MOTIF.
FT DOMAIN 313 334 LEUCINE-ZIPPER.
SQ SEQUENCE 370 AA; 38655 MW; 76A9517EF9C777C CRC64;
```

```
FT DOMAIN 312 333 LEUCINE-ZIPPER.
SQ SEQUENCE 369 AA; 38457 MW; 288E464708DAC7D CRC64;

Query Match          90.9%; Score 30; DB 1; Length 369;
Best Local Similarity 83.3%; Pred. No. 2.3e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCTSG 6
DB 232 GGGTAG 237

RESULT 28
MAF2_MOUSE        STANDARD; PRT; 370 AA.
ID MAF2_MOUSE
AC P54843;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE TRANSCRIPTION FACTOR MAF2 (PROTO-ONCOGENE C-MAF).
GN MAF2 OR MAF.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=Cerebellum;
RX MEDLINE=95097997; PubMed=7799931;
RA Kirschner C., Morgan J.L.;
RT *The maf proto-oncogene stimulates transcription from multiple sites
RT in a promoter that directs Purkinje neuron-specific gene
RT expression.
RL Mol. Cell. Biol. 15:246-254(1995).
CC -1- FUNCTION: THE C-MAF INTERACTION SITE WAS MAPPED TO THE SEQUENCE
CC 5'-[GTG]G[CG]N[GT]NCTAGNN-3' IN THE L7 PROMOTER. IT MAY INTERACT
CC WITH ADDITIONAL BASIC-ZIPPER PROTEINS THAT DETERMINE A SUBTYPE OF
CC MAF-RESPONSIVE ELEMENT BINDING.
CC -1- SUBUNIT: HOMO- OR HETERODIMER.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: BELONGS TO THE BZIP FAMILY. MAF SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: S74567; AAB32820.1; -.
DR HSSP: P05412; IJUN.
DR TRANSFAC: T01432; -.
DR MGD: MGI:96909; Maf.
DR InterPro: IPR001871; bZIP.
DR SMART: SM00338; BRLZ; 1.
KW Proto-oncogene; Transcription regulation; DNA-binding; Activator;
KW Nuclear protein.
FT DOMAIN 139 146 POLY-ALA.
FT DOMAIN 169 173 POLY-ALA.
FT DOMAIN 180 187 POLY-HIS.
FT DOMAIN 191 194 POLY-HIS.
FT DOMAIN 207 210 POLY-SER.
FT DOMAIN 212 220 POLY-GLY.
FT DOMAIN 225 235 POLY-GLY.
FT DNA_BIND 285 311 BASIC MOTIF.
FT DOMAIN 313 334 LEUCINE-ZIPPER.
SQ SEQUENCE 370 AA; 38655 MW; 76A9517EF9C777C CRC64;

Query Match          90.9%; Score 30; DB 1; Length 370;
Best Local Similarity 83.3%; Pred. No. 2.3e+02;
```

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GCGTSG 6  
 DB 233 GCGTAG 238

RESULT 29

DNA\_BACST STANDARD: PRT: 372 AA.  
 AC P17557:  
 DT 01-AUG-1990 (rel. 15, Created)  
 DT 01-AUG-1990 (rel. 15, Last sequence update)  
 DT 01-OCT-1994 (rel. 30, Last annotation update)  
 DE ALANINE DEHYDROGENASE (EC 1.4.1.1).  
 OS Bacillus stearothermophilus.  
 OC Bacteria: Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Geobacillus.  
 OX NCBI\_TaxID=1422;  
 RN [1]  
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.  
 RC STRAIN-IFO 12550;  
 RX MEDLINE=90254112; PubMed=2340214;  
 RA Kuroda S., Tanizawa K., Sakamoto Y., Tanaka H., Soda K.;  
 RT "Alanine dehydrogenases from two Bacillus species with distinct  
 RT thermostabilities: molecular cloning, DNA and protein sequence  
 RT determination, and structural comparison with other  
 RT NAD(P)(+)-dependent dehydrogenases.";  
 RT Biochemistry 29:1009-1015(1990).  
 CC -1- FUNCTION: THIS ENZYME IS A KEY FACTOR IN THE ASSIMILATION OF  
 CC L-ALANINE AS AN ENERGY SOURCE THROUGH THE TRICARBOXYLIC ACID  
 CC CYCLE DURING SPOULATION.  
 CC -1- CATALYTIC ACTIVITY: L-ALANINE + H(2)O + NAD(+) - PYRUVATE +  
 CC NH(3) + NADH.  
 CC -1- SUBUNIT: HOMOHexamER.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -1- MISCELLANEOUS: THE B.STEAROTHERMOPHILUS ENZYME RETAINED ABOUT 50%  
 CC OF ITS INITIAL ACTIVITY WHEN HEATED AT 85 DEGREES CELSIUS FOR 5  
 CC MIN AT PH 7.2, WHEREAS THE B.SPHERICUS ENZYME LOST THE SAME  
 CC ACTIVITY WHEN HEATED AT ONLY 65 DEGREES CELSIUS FOR 5 MIN.  
 CC -1- SIMILARITY: STRONG, TO OTHER ALANINE DEHYDROGENASES AND PARTIAL  
 CC WITH PYRIDINE NUCLEOTIDE TRANSFERASES.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: M33299; AAA22211.1; -;  
 DR PIR: B34261; B34261.  
 DR InterPro: IPR0000598; NAD\_binding.  
 DR InterPro: IPR000205; NAD\_binding.  
 DR Pfam: PF01262; AladH\_PNT; 1.  
 DR PROSITE: PS00836; AladH\_PNT\_1; 1.  
 DR PROSITE: PS00837; AladH\_PNT\_2; 1.  
 DR Oxidoreductase; NAD; Sporulation.  
 FT ACT SITE 96 96 POTENTIAL.  
 FT NP\_BIND 170 200 NAD (BY SIMILARITY).  
 SO SEQUENCE 372 AA; 39694 MW; 99A8EB621ED1F23A CRC64;

Query Match 90.9%; Score 30; DB 1; Length 372;  
 Best Local Similarity 83.3%; Pred. No. 2.3e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGTSG 6  
 DB 175 GCGTAG 180

RESULT 30  
 LMXA\_MESAU STANDARD: PRT: 382 AA.  
 AC Q04650;  
 DT 01-JUN-1994 (rel. 29, Created)  
 DT 01-JUN-1994 (rel. 29, Last sequence update)  
 DT 30-MAY-2000 (rel. 39, Last annotation update)  
 DE LIM/HOMEOBOX PROTEIN LMXA (LMX-1) (LIM-HOMEOBOX PROTEIN 1).  
 GN LMXA OR LMX1.  
 OS Mesocricetus auratus (Golden hamster).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 OC Mesocricetus  
 OX NCBI\_TaxID=10036;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Pancreas;  
 RX MEDLINE=93051335; PubMed=1358758;  
 RA German M.S., Wang J., Chadwick R.B., Rutter W.J.;  
 RT "Synergistic activation of the insulin gene by a LIM-homeo domain  
 RT protein and a basic helix-loop-helix protein: building a functional  
 RT insulin minihancer complex";  
 RT Genes Dev. 6:2165-2176(1992).  
 CC -1- FUNCTION: ACTS AS A TRANSCRIPTIONAL ACTIVATOR BY BINDING TO THE  
 CC FLAT ELEMENT WHICH IS A BETA-CELL-SPECIFIC TRANSCRIPTIONAL  
 CC ENHANCER FOUND IN THE RAT INSULIN GENE AND INTERACTING THROUGH  
 CC ITS LIM-CONTAINING AMINO TERMINUS WITH SH-PAN, A HELIX-LOOP-HELIX  
 CC PROTEIN.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- SIMILARITY: TO OTHER HOMEOBOX DOMAINS. BELONGS TO THE LIM  
 CC SUBFAMILY.  
 CC -1- SIMILARITY: CONTAINS 2 LIM DOMAINS. THE LIM DOMAIN BINDS 2 ZINC  
 CC IONS.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: X81406; CAA57163.1; -;  
 DR PIR: B46233; B46233.  
 DR HSSP: P02836; 1HDD.  
 DR TRANSFAC: T01958;  
 DR InterPro: IPR001356; Homeobox.  
 DR InterPro: IPR001781; LIM.  
 DR Pfam: PF00046; homeobox; 1.  
 DR ProDom: PD000094; LIM; 2.  
 DR SMART: SM00389; HOX; 1.  
 DR SMART: SM00132; LIM; 2.  
 DR PROSITE: PS00478; LIM\_DOMAIN\_1; 2.  
 DR PROSITE: PS50023; LIM\_DOMAIN\_2; 2.  
 DR PROSITE: PS00027; HOMEOBOX\_1; 1.  
 DR PROSITE: PS50071; HOMEOBOX\_2; 1.  
 DR Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
 KW LIM domain; Repeat; Metal-binding; Zinc; Activator;  
 KW Transcription regulation.  
 FT DOMAIN 35 85 LIM 1.  
 FT DOMAIN 94 147 LIM 2.  
 FT DNA\_BIND 195 254 HOMEOBOX.  
 FT DOMAIN 257 265 POLY-GLN.  
 SO SEQUENCE 382 AA; 42802 MW; 2675B6F298BE9486 CRC64;

Query Match 90.9%; Score 30; DB 1; Length 382;  
 Best Local Similarity 83.3%; Pred. No. 2.4e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGTSG 6  
 DB 175 GCGTAG 180

Db 277 GCGTAG 282

## RESULT 31

SERC\_YEAST STANDARD; PRT; 395 AA.

AC P33330;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DE 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE PHOSPHOSERINE AMINOTRANSFERASE (EC 2.6.1.52) (PSATF).  
 GN SERL OR SERC OR YOR184W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;

RA MEDLINE=94287715; PubMed=8017107;  
 RA Belhuneur P., Fortin N., Clark M.W.;  
 RT "A gene from Saccharomyces cerevisiae which codes for a protein with  
 RT significant homology to the bacterial 3-phosphoserine  
 RT aminotransferase."  
 RL Yeast 10:385-389(1994).

RP SEQUENCE FROM N.A.  
 RA Mocher K., Kuenzler M., Braus G., Entian K.D.;  
 RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.

RA Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
 RL Hughes B., Pohl T.M.;  
 CC -1- CATALYTIC ACTIVITY: O-PHOSPHO-L-SERINE + 2 OXOGUTARATE -

CC -1- 3-PHOSPHONOOXYPRUVATE + L-GLUTAMATE.  
 CC -1- COPACITOR: PYRIDOXAL PHOSPHATE.  
 CC -1- PATHWAY: REQUIRED BOTH IN MAJOR PHOSPHORYLATED PATHWAY OF SERINE  
 CC BIOSYNTHESIS AND IN THE BIOSYNTHESIS OF PYRIDOXINE.  
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO CLASS-V OF PYRIDOXAL-PHOSPHATE-DEPENDENT  
 CC AMINOTRANSFERASES.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: L20917; AAA20886.1; -;  
 DR EMBL: U19714; AAA85703.1; -;  
 DR EMBL: 275092; CAA93933.1; -;  
 DR PIR: S42680; S42680.  
 DR HSSP: P23721; 1BUN.  
 DR SWISS-2DPAGE: P33330; YEAST.  
 DR YEPD: 5420; -;  
 DR SGD: S0005710; SERL.  
 DR InterPro: IPR000192; AminoTransf\_Class\_V.  
 DR InterPro: IPR003248; Phosphoser\_AminoTransf.  
 DR Pfam: PF00266; aminoTran\_5; 1.  
 DR ProDom: PD001544; Phosphoser\_AminoTransf; 1.  
 DR PROSITE: PS00595; AA\_TRANSFERS\_CLASS\_5; 1.  
 KW Serine biosynthesis; Transferase; AminoTransferase;  
 KW Pyridoxal phosphate  
 FT BINDING 218 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 FT CONFLICT 258 FD -> LH (IN REF. 2).  
 SQ SEQUENCE 395 AA; 43415 MW; A22F020CC69B4BC CRC64;

Query Match 90.9%; Score 30; DB 1; Length 395;  
 Best Local Similarity 83.3%; Pred. No. 2.5e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGTSG 6  
Db 78 GCGTTG 83

## RESULT 32

DDL\_TREPA STANDARD; PRT; 396 AA.

AC 083676;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DE 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE D-ALANINE--D-ALANINE LIGASE (EC 6.3.2.4) (D-ALANYLALANINE SYNTHETASE)  
 DE (D-ALA-D-ALA LIGASE).  
 GN DDL OR DDLA OR TP0670.  
 OS Treponema pallidum.  
 OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.

OX NCBI\_TaxID=160;  
 RA SEQUENCE FROM N.A.  
 RC STRAIN=NICHOLES;  
 RX MEDLINE=98332770; PubMed=9665876;

RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,  
 RA Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,  
 RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,  
 RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Uitterback T.,  
 RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,  
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,  
 RA Venter J.C.;  
 RT "Complete genome sequence of Treponema pallidum, the syphilis  
 RT spirochete."  
 RL Science 281:375-388(1998).

CC -1- FUNCTION: CELL WALL FORMATION (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: ATP + D-ALANINE + D-ALANINE -> ADP +  
 CC ORTHOPHOSPHATE + D-ALANYL-D-ALANINE.  
 CC -1- PATHWAY: ALONG WITH ALANINE RACEMASE, IT MAKES UP THE D-ALANINE  
 CC BRANCH OF THE PEPTIDOGLYCAN BIOSYNTHETIC ROUTE (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE D-ALANINE--D-ALANINE LIGASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AE001241; AAC26568.1; -;  
 DR TIGR: TP0670; -;  
 DR InterPro: IPR000291; Dala\_Data\_Ligase.  
 DR Pfam: PF01820; Dala\_Data\_Ligase; 1.  
 DR PROSITE: PS00843; DALA\_DATA\_LIGASE\_1; 1.  
 DR PROSITE: PS00844; DALA\_DATA\_LIGASE\_2; 1.  
 KW Ligase; Cell wall; Peptidoglycan synthesis; Complete proteome.  
 SQ SEQUENCE 396 AA; 42728 MW; FB5829335B6111A5 CRC64;

Query Match 90.9%; Score 30; DB 1; Length 396;  
 Best Local Similarity 83.3%; Pred. No. 2.5e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGTSG 6  
Db 79 GCGTAG 84

## RESULT 33

MURG\_MYCLE STANDARD; PRT; 407 AA.

AC 069552;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)



DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE UDP-N-ACETYLGLUCOSAMINE--N-ACETYLURAMYL--(PENTAPEPTIDE)  
 DE PYROPHOSPHORYL-UNDECAPRENOL-N-ACETYLGLUCOSAMINE TRANSFERASE  
 DE (EC 2.4.1.-) (UNDECAPRENYL-PP-MURNAC-PENTAPEPTIDE-UDPGLCNAC GLCNAC  
 TRANSFERASE)  
 GN MURG OR ML0914 OR MLCB266.02C.  
 OS Mycobacterium leprae.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1769;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-TN;  
 RX MEDLINE=21128732; PubMed=11234002;  
 RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,  
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,  
 RA Holroyd S., Hornsby T., Jagsels K., Lacroix C., Maclean J., Moule S.,  
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.-A., Rutherford K.M.,  
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
 RA Barrell B.G.;  
 RT "Massive gene decay in the leprosy bacillus.";  
 RL Nature 409:1007-1011(2001).  
 CC -1- FUNCTION: CELL WALL FORMATION. CATALYZES THE TRANSFER OF A GLCNAC  
 SUBUNIT ON UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-PENTAPEPTIDE (LIPID  
 INTERMEDIATE I) TO FORM UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-  
 (PENTAPEPTIDE)GLCNAC (LIPID INTERMEDIATE II) (BY SIMILARITY).  
 CC -1- PATHWAY: LAST STEP OF PEPTIDOGLYCAN BIOSYNTHESIS.  
 CC -1- SUBCELLULAR LOCATION: MEMBRANE ASSOCIATED (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE MURG FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: AL022602; CNA18668.1; -  
 DR EMBL: AL583920; CAC31295.1; -  
 DR Leproma: ML0914; -  
 KM Transferase: Glycosyltransferase; Cell division; Cell wall; Membrane;  
 KW Peptidoglycan synthesis; Complete proteome.  
 SQ SEQUENCE 407 AA; 42422 MW; B757B73C092C53F9 CRC64;

Query Match 90.9%; Score 30; DB 1; Length 407;  
 Best Local Similarity 83.3%; Pred. No. 2.5e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGTSG 6  
 ||||:|  
 DB 35 GGGTAG 40

RESULT 34  
 MURG\_MYCTU STANDARD; PRT; 410 AA.  
 AC 006224;  
 ID 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE UDP-N-ACETYLGLUCOSAMINE--N-ACETYLURAMYL--(PENTAPEPTIDE)  
 DE PYROPHOSPHORYL-UNDECAPRENOL-N-ACETYLGLUCOSAMINE TRANSFERASE  
 DE (EC 2.4.1.-) (UNDECAPRENYL-PP-MURNAC-PENTAPEPTIDE-UDPGLCNAC GLCNAC  
 TRANSFERASE)  
 GN MURG OR RV2153C OR MT2212 OR MTC2720.15.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.

OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badoock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagsels K., Krogh A., Mclean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.,  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 CC [2]  
 CC SEQUENCE FROM N.A.  
 RC STRAIN-CDC 1551 / Oghkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Hatt D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 laboratory strains.";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: CELL WALL FORMATION. CATALYZES THE TRANSFER OF A GLCNAC  
 SUBUNIT ON UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-PENTAPEPTIDE (LIPID  
 INTERMEDIATE I) TO FORM UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-  
 (PENTAPEPTIDE)GLCNAC (LIPID INTERMEDIATE II) (BY SIMILARITY).  
 CC -1- PATHWAY: LAST STEP OF PEPTIDOGLYCAN BIOSYNTHESIS.  
 CC -1- SUBCELLULAR LOCATION: MEMBRANE ASSOCIATED (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE MURG FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: Z95388; CAB08640.1; -  
 DR EMBL: AE007068; AAK46496.1; -  
 DR TIGR: MT2212; -  
 DR TubercuList: RV2153C; -  
 KM Transferase: Glycosyltransferase; Cell division; Cell wall; Membrane;  
 KW Peptidoglycan synthesis; Complete proteome.  
 SQ SEQUENCE 410 AA; 41860 MW; 96902AFE356FC30B CRC64;

Query Match 90.9%; Score 30; DB 1; Length 410;  
 Best Local Similarity 83.3%; Pred. No. 2.6e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGTSG 6  
 ||||:|  
 DB 42 GGGTAG 47

RESULT 35  
 YA68\_MYCTU STANDARD; PRT; 463 AA.  
 AC 053416;  
 ID 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE HYPOTHETICAL PE-PCRS FAMILY PROTEIN RV1068C.  
 DE RV1068C OR MT1097 OR MTV017.21C.  
 GN Mycobacterium tuberculosis.  
 OS Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.



DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE (R)-MANDELONITRILE LYASE ISOFORM 2 PRECURSOR (EC 4.1.2.10)  
 DE (HYDROXYNITRILE LYASE 2) ((R)-OXYNITRILE 2).  
 GN MDL2.  
 OS Prunus serotina (black cherry).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids I; Rosales; Rosaceae; Prunus.  
 OX NCBI\_TaxID=23207;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Hu Z., Poulton J.E.;  
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: INVOLVED IN CYANOGENS, THE RELEASE OF HCN FROM INJURED  
 CC TISSUES. CATALYSES THE STEREOSPECIFIC ADDITION OF HCN TO A VARIETY  
 CC OF ALDEHYDES IN VITRO. IS A MAJOR SEED CONSTITUENT, AND COULD HAVE  
 CC THE ADDITIONAL ROLE OF A STORAGE FORM FOR REDUCED NITROGEN (BY  
 CC SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: MANDELONITRILE + CYANIDE + BENZALDEHYDE.  
 CC -1- COFACTOR: FAD.  
 CC -1- SUBUNIT: MONOMER.  
 CC -1- SUBCELLULAR LOCATION: PRIMARYLY FOUND WITHIN PROTEIN BODIES OF THE  
 CC CYTOPLASMIC PARENCHYMA CELLS WITH LESSER AMOUNTS WITHIN THE  
 CC PROCAMBIUM (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE GMC OXIDOREDUCTASES FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: AF040078; AAB96763.1; -  
 CC EMBL: AF040079; AAB96764.1; -  
 CC InterPro: IPR00172; GMC\_Oxred.1.  
 DR Pfam: PF00732; GMC\_Oxred.1.  
 DR PROSITE: PS00623; GMC\_OXRED.1; 1.  
 DR PROSITE: PS00624; GMC\_OXRED.2; 1.  
 KW Lyase; Glycoprotein; FAD; Flavoprotein; Signal; Multigene family.  
 FT SIGNAL 1 28  
 FT CHAIN 1 576  
 FT NP\_BIND 29 576  
 FT CARBOHYD 31 83  
 FT CARBOHYD 47 47  
 FT CARBOHYD 76 76  
 FT CARBOHYD 146 146  
 FT CARBOHYD 151 151  
 FT CARBOHYD 163 163  
 FT CARBOHYD 184 184  
 FT CARBOHYD 219 219  
 FT CARBOHYD 268 268  
 FT CARBOHYD 310 310  
 FT CARBOHYD 381 381  
 FT CARBOHYD 413 413  
 FT CARBOHYD 421 421  
 FT CARBOHYD 468 468  
 SQ SEQUENCE 576 AA: 62724 MW: 570DC7853AEAD3EA CRC64;

Query Match 90.9%; Score 30; DB 1; Length 576;  
 Best Local Similarity 83.3%; Pred. No. 3.6e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGGTSG 6  
 |||||  
 DB 61 GGGTAG 66

RESULT 38  
 CM70\_HUMAN

ID OM70\_HUMAN STANDARD: PRT: 608 AA.  
 AC 094826;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE MITOCHONDRIAL PRECURSOR PROTEIN IMPORT RECEPTOR (TRANSLOCASE OF OUTER  
 DE MEMBRANE TOM70).  
 GN TOM70A OR KIAA0719.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Tissue-Brain;  
 RC MEDLINE-99087487; PubMed-9872452;  
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,  
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RT for large proteins in vitro.  
 RL DNA Res. 5:277-286(1998).  
 CC -1- FUNCTION: RECEPTOR THAT ACCELERATES THE IMPORT OF ALL  
 CC MITOCHONDRIAL PRECURSOR PROTEINS (BY SIMILARITY).  
 CC -1- SUBUNIT: FORMS PART OF THE RECEPTOR COMPLEX THAT CONSISTS OF AT  
 CC LEAST 8 DIFFERENT PROTEINS (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL  
 CC -1- SIMILARITY: CONTAINS 10 TPR REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: AB018262; BAA34439.1; -  
 CC HSSP: P53041; 1A17.  
 DR InterPro: IPR01440; TPR.  
 DR Pfam: PF00515; TPR; 10.  
 DR SMART: SM00028; TPR; 8.  
 KW Mitochondrion; Outer membrane; Transmembrane; Repeat; TPR repeat.  
 FT DOMAIN 1 8  
 FT TRANSMEM 9 29  
 FT DOMAIN 30 608  
 FT REPEAT 114 147  
 FT REPEAT 153 186  
 FT REPEAT 294 327  
 FT REPEAT 329 362  
 FT REPEAT 367 400  
 FT REPEAT 401 434  
 FT REPEAT 440 475  
 FT REPEAT 476 509  
 FT REPEAT 511 544  
 FT REPEAT 545 578  
 SQ SEQUENCE 608 AA: 67454 MW: 5A9FCAAA8582480 CRC64;

Query Match 90.9%; Score 30; DB 1; Length 608;  
 Best Local Similarity 83.3%; Pred. No. 3.8e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGGTSG 6  
 |||||  
 DB 25 GGGTAG 30

RESULT 39  
 OM70\_MOUSE STANDARD: PRT: 611 AA.  
 ID OM70\_MOUSE  
 AC 09C2W5;

DT 20-AUG-2001 (Rel. 40, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last sequence update)  
DE MITOCHONDRIAL PRECURSOR PROTEIN IMPORT RECEPTOR (TRANSLOCASE OF OUTER  
MEMBRANE TOM70).  
GN TOM70A.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Embryo;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Akawa T., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
Atakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
Aizawa K., Izeva M., Nishi K., Kiyosawa H., Kondo S., Yamamaki I.,  
Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
R. Fleischmann W., Gaasterland T., Glasl C., King B., Koehliwa H.,  
Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
Schramm L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,  
Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,  
Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
Gustincic S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,  
Suzuki H., Toyooka K., Wang K.H., Wenzel K., Whitaker C., Wilming L.,  
Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
CC -1- FUNCTION: RECEPTOR THAT ACCELERATES THE IMPORT OF ALL  
MITOCHONDRIAL PRECURSOR PROTEINS (BY SIMILARITY).  
CC -1- SUBUNIT: FORMS PART OF THE RECEPTOR COMPLEX THAT CONSISTS OF AT  
LEAST 8 DIFFERENT PROTEINS (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL  
OUTER MEMBRANE (BY SIMILARITY).  
CC -1- SIMILARITY: CONTAINS 10 TPR REPEATS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AK012084; BAB28018.1;  
DR MGD: MGI:106295; D16msu109e.  
DR InterPro: IPR001440; TPR.  
DR Pfam: PF00515; TPR. 10.  
DR SMART: SM00028; TPR. 10.  
KW Mitochondrion; Outer membrane; Transmembrane; Repeat; TPR repeat.  
FT DOMAIN 1 30 INTERMEMBRANE (POTENTIAL).  
FT TRANSMEM 1 30 POTENTIAL.  
FT DOMAIN 31 611 CYTOPLASMIC (POTENTIAL).  
FT REPEAT 117 150 TPR 1.  
FT REPEAT 156 189 TPR 2.  
FT REPEAT 297 330 TPR 3.  
FT REPEAT 332 365 TPR 4.  
FT REPEAT 370 403 TPR 5.  
FT REPEAT 404 437 TPR 6.  
FT REPEAT 445 478 TPR 7.  
FT REPEAT 479 512 TPR 8.  
FT REPEAT 514 547 TPR 9.  
FT REPEAT 548 581 TPR 10.  
SQ SEQUENCE 611 AA; 67521 MW; 486FB79FC4C5B4F CRC64;

Best Local Similarity 83.3%; Pred. No. 3.8e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGTSG 6  
||||:|  
Db 26 GGTAG 31

RESULT 40

UVRAL THEME STANDARD; PRT; 916 AA.

AC 09WYV0;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE EXCINUCLEASE ABC SUBUNIT A.  
GN UVRAL OR TM0480.  
OS Thermotoga maritima.  
OC Bacteria; Thermotogales; Thermotoga.  
OX NCBI\_TaxID=2336;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MSB8 / DSM 3109;  
RX MEDLINE=99287316; PubMed=10360571;  
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,  
Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,  
Heidelberg J., Sutton H.O., Venter J.C., Fraser C.M.;  
RT "Evidence for lateral gene transfer between Archaea and Bacteria from  
genome sequence of Thermotoga maritima.";  
RL Nature 399:323-329(1999).  
CC -1- FUNCTION: THE ABC EXCISION NUCLEASE IS A DNA REPAIR ENZYME THAT  
CATALYZES THE EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS  
PRODUCING OLIGOMERS HAVING THE MODIFIED BASE(S). UVRAL IS AN ATPASE  
AND A DNA-BINDING PROTEIN THAT PREFERENTIALLY BINDS SINGLE-  
STRANDED OR UV-IRRADIATED DOUBLE-STRANDED DNA (BY SIMILARITY).  
CC -1- SUBUNIT: CONSISTS OF THREE SUBUNITS: UVRAL, UVRB AND UVRC.  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY  
(ABC TRANSPORTERS). CONTAINS TWO ABC DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AE001725; AAD35564.1;  
DR TIGR: TM0480;  
DR InterPro: IPR003439; ABC\_transportr.  
DR Pfam: PF00005; ABC\_tran; 2.  
DR PROSITE: PS00211; ABC\_TRANSPORTER; 2.  
KW SOS response; Excision nuclease; DNA repair; ATP-binding; Repeat;  
FM DNA-binding; Zinc-finger; Complete proteome.  
FT NP\_BIND 31 38 ATP (POTENTIAL).  
FT NP\_BIND 617 624 ATP (POTENTIAL).  
FT ZN\_FING 252 279 C4-TYPE.  
FT ZN\_FING 716 742 C4-TYPE.  
SQ SEQUENCE 916 AA; 102967 MW; 54FBA620783791A CRC64;

Query Match 90.9%; Score 30; DB 1; Length 916;  
Best Local Similarity 83.3%; Pred. No. 5.6e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGTSG 6  
||||:|  
Db 541 GGTGNG 546

Mon Feb 4 08:21:03 2002

us-09-642-660-10.rsp

Page 21

Search completed: February 4, 2002, 08:01:10  
Job time: 54 sec

---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 4, 2002, 08:00:16 ; Search time 35.81 Seconds  
(without alignments)  
24.508 Million cell updates/sec

Title: us-09-642-660-10-  
Perfect score: 33  
Sequence: 1 GCGTSG 6

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 100 summaries

Database :

1: sptrembl\_17:\*  
2: sp-archaea:\*  
3: sp-bacteria:\*  
4: sp-fungi:\*  
5: sp-invertebrate:\*  
6: sp-mammal:\*  
7: sp-mhc:\*  
8: sp-organellae:\*  
9: sp-phage:\*  
10: sp-plant:\*  
11: sp-rodent:\*  
12: sp-virus:\*  
13: sp-vertebrate:\*  
14: sp-unclassified:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	100.0	218	10	09FJ02
2	33	100.0	221	12	09DW22
3	33	100.0	224	3	09HPT3
4	33	100.0	251	2	09EV39
5	33	100.0	256	4	09NZ70
6	33	100.0	256	4	09NP55
7	33	100.0	275	5	09VYLO
8	33	100.0	279	5	062416
9	33	100.0	289	10	080450
10	33	100.0	303	10	09LTP4
11	33	100.0	315	5	09YIH0
12	33	100.0	348	4	075953
13	33	100.0	427	8	09GE43
14	33	100.0	431	8	09TWM3
15	33	100.0	449	8	003112
16	33	100.0	451	8	031681
17	33	100.0	468	8	09BCR9
18	33	100.0	469	8	09GHW8
19	33	100.0	470	4	09H1J0

20	33	100.0	475	8	09NM74	09cm74 rhynchosleg
21	33	100.0	475	8	09TM63	09cm63 platyhyphid
22	33	100.0	483	4	09B217	09bz17 homo sapien
23	33	100.0	484	5	09VSE9	09vse9 drosophila
24	33	100.0	503	2	09A4M2	09a4m2 caulobacter
25	33	100.0	504	2	09AA21	09aa21 caulobacter
26	33	100.0	504	5	027403	027403 drosophila
27	33	100.0	504	5	09VLAS	09vla5 drosophila
28	33	100.0	528	5	09W0N5	09w0n5 drosophila
29	33	100.0	559	10	082435	082435 prunus sero
30	33	100.0	563	10	09FJ97	09fj97 arabisdopsis
31	33	100.0	565	2	099Y17	099y17 streptococc
32	33	100.0	570	2	099S77	099s77 staplylococ
33	33	100.0	574	10	082784	082784 prunus sero
34	33	100.0	577	3	P94865	P94865 lactobacill
35	33	100.0	578	3	09P6V0	09p6v0 neurospora
36	33	100.0	618	10	09SIN1	09sin1 arabisdopsis
37	33	100.0	668	12	056765	056765 human cytom
38	33	100.0	691	10	09ASR6	09asr6 arabisdopsis
39	33	100.0	809	11	09BRT4	09btr4 mus musculu
40	33	100.0	817	5	09BJL1	09bjl1 caenorhabdi
41	33	100.0	856	5	09VL95	09v195 drosophila
42	33	100.0	856	5	09UGA0	09ugao drosophila
43	33	100.0	1096	5	017368	017368 caenorhabdi
44	33	100.0	1144	5	09TVG6	09tv6 caenorhabdi
45	33	100.0	1164	5	09V161	09v161 drosophila
46	33	100.0	1192	5	017346	017346 caenorhabdi
47	33	100.0	1192	5	09TW45	09tw45 caenorhabdi
48	33	100.0	1268	10	09SDC5	09sdc5 oryza sativ
49	33	100.0	1408	10	09FMU0	09fmu0 oryza sativ
50	33	100.0	2071	3	09HF75	09hf75 ashbya goss
51	33	100.0	2158	4	09Y6T3	09y6t3 homo sapien
52	33	100.0	2265	4	09Y6T3	09y6t3 homo sapien
53	33	100.0	130	2	09PD43	09pd43 xylella fas
54	33	100.0	143	10	09SMU5	09smu5 arabisdopsis
55	33	100.0	185	1	09HS10	09hs10 halobacteri
56	33	100.0	215	3	043117	043117 collettortic
57	33	100.0	234	10	09AWI3	09awi3 oryza sativ
58	33	100.0	254	10	024568	024568 zea mays (m
59	33	100.0	258	4	09H802	09h802 homo sapien
60	33	100.0	258	4	09BUD2	09bud2 homo sapien
61	33	100.0	276	2	09RS95	09rs95 delnoccocus
62	33	100.0	293	10	09AVG2	09avg2 oryza sativ
63	33	100.0	310	2	008454	008454 amycolata s
64	33	100.0	315	10	09ATN8	09atn8 pennisetum
65	33	100.0	340	5	017799	017799 caenorhabdi
66	33	100.0	346	10	09ZSD4	09zsd4 arabisdopsis
67	33	100.0	358	4	09NMD0	09nmd0 homo sapien
68	33	100.0	371	2	09CEV6	09cev6 mycobacteri
69	33	100.0	372	2	09LAH3	09lah3 brevbacter
70	33	100.0	373	5	09XYH2	09xyh2 toxoplasma
71	33	100.0	382	2	09EUS5	09eus5 streptomyce
72	33	100.0	382	10	09MIW9	09miw9 arabisdopsis
73	33	100.0	384	2	035028	035028 bacillus su
74	33	100.0	386	4	09NMN3	09nmn3 homo sapien
75	33	100.0	392	2	09Z4J4	09z4j4 ervinia chr
76	33	100.0	397	2	099XC4	099xc4 staphylococ
77	33	100.0	402	2	09ESN3	09esn3 bacillus ce
78	33	100.0	430	5	09VNB6	09vnb6 drosophila
79	33	100.0	467	2	09JZ00	09jz00 nelisseria m
80	33	100.0	467	2	09JUT5	09jut5 nelisseria m
81	33	100.0	473	5	099TX5	099tx5 staphylococ
82	33	100.0	475	5	09N6R7	09n6r7 drosophila
83	33	100.0	475	5	09N6R7	09n6r7 drosophila
84	33	100.0	480	10	09C753	09c753 arabisdopsis
85	33	100.0	481	5	09GUT7	09gut7 diatrobica
86	33	100.0	484	2	09XNU2	09xnu2 vibrio chol
87	33	100.0	497	10	09SA43	09sa43 arabisdopsis
88	33	100.0	499	2	09A9H8	09a9h8 caulobacter
89	33	100.0	501	10	09X168	09x168 arabisdopsis
90	33	100.0	502	2	09A4M5	09a4m5 caulobacter
91	33	100.0	503	10	09X169	09x169 arabisdopsis
92	33	100.0	504	2	09A607	09a607 caulobacter

93 30 90.9 509 2 0944M3 0944M3 caulobacter  
 94 30 90.9 510 2 09K3J1 09K3J1 streptomyc  
 95 30 90.9 510 2 09AK9 09AK9 caulobacter  
 96 30 90.9 515 5 09U5I8 09U5I8 drosophila  
 97 30 90.9 524 12 09IWC1 09IWC1 phococena sp  
 98 30 90.9 527 2 067206 067206 aquifex aeo  
 99 30 90.9 537 2 09RPG7 09RPG7 burkholderi  
 100 30 90.9 538 2 P95480 P95480 pseudomonas

## ALIGNMENTS

RESULT 1

O9FJ02 PRELIMINARY: PRT; 218 AA.

ID O9FJ02  
 AC O9FJ02;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DE GENOMIC DNA, CHROMOSOME 5, TAC CLONE:K22F20.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_Taxid=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-COLUMBIA;  
 RX MEDLINE=99087489; PubMed=9872454;  
 RA Nakamura Y., Sato S., Asamizu E., Kaneo T., Kotani H., Miyajima N.,  
 RA Tabata S.;  
 RT \*Structural analysis of Arabidopsis thaliana chromosome 5. VII.  
 RT Sequence features of the regions of 1,013,767 bp covered by sixteen  
 RT physically assigned P1 and TAC clones.\*;  
 RL DNA Res. 5:297-308(1996).  
 DR EMBL: AB016873; BAB10357.1; -;  
 SQ SEQUENCE 218 AA; 22937 MW; C8551D884AB1FDD5 CRC64;

Query Match 100.0%; Score 33; DB 10; Length 218;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
 |||||  
 DB 78 GCGTSG 83

RESULT 2

O9DM22 PRELIMINARY: PRT; 221 AA.

ID O9DM22  
 AC O9DM22;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DE PRI71.1.  
 CN R171.1.  
 OS Rat cytomegalovirus (strain Mastricht).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Betaherpesvirinae; Muromegalovirus.  
 OX NCBI\_Taxid=79700;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MASTRICHT;  
 RX MEDLINE=96357047; PubMed=8764031;  
 RA Vink C., Beuken E., Bruggeman C.A.;  
 RT \*Structure of the rat cytomegalovirus genome termint.;  
 RL J. Virol. 70:5221-5229(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MASTRICHT;  
 RX MEDLINE=20366325; PubMed=10906222;

RA Vink C., Beuken E., Bruggeman C.A.;  
 RT \*Complete DNA sequence of the rat cytomegalovirus genome.\*;  
 RL J. Virol. 74:7656-7665(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MASTRICHT;  
 RX MEDLINE=20473137; PubMed=11018281;  
 RA Gruljthuisen Y.K., Beuken E., Bruggeman C.A., Vink C.;  
 RT \*Rat cytomegalovirus R89 is a highly conserved gene which expresses a  
 RT spliced transcript.\*;  
 RL Virus Res. 69:119-130(2000).  
 DR EMBL: AF232689; AAF99268.1; -;  
 SQ SEQUENCE 221 AA; 24389 MW; CD36BC51A9AB69B3 CRC64;

Query Match 100.0%; Score 33; DB 12; Length 221;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
 |||||  
 DB 184 GCGTSG 189

RESULT 3

O9HFT3 PRELIMINARY: PRT; 224 AA.

ID O9HFT3  
 AC O9HFT3;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DE XYLANASE (FRAGMENT).  
 OS Pliomycetes communis.  
 OC Eukaryota; Fungi; Chytridiomycota; Neocallimastixales;  
 OC Neocallimastixaceae; Pliomycetes.  
 OX NCBI\_Taxid=137782;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kam D.K., Kobayashi Y., Ha J.K.;  
 RT \*Transformation of Pliomycetes communis xylanase to B. fibrisolvens.\*;  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF297649; ANG18439.1; -;  
 DR Interpro: IPR001137; Glyco\_hydro\_11.  
 DR Pfam: PF00457; Glyco\_hydro\_11; 1.  
 DR PRINTS: PR00911; GLHYDRLASE11.  
 FT NON\_TER 1  
 FT 1  
 SQ SEQUENCE 224 AA; 24976 MW; 93E512C9DF176DDE CRC64;

Query Match 100.0%; Score 33; DB 3; Length 224;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
 |||||  
 DB 207 GCGTSG 212

RESULT 4

O9EV39 PRELIMINARY: PRT; 251 AA.

ID O9EV39  
 AC O9EV39;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DE PHOSPHOTRANSFERASE LACE (FRAGMENT).  
 CN LACE.  
 OS Streptococcus gordonii.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_Taxid=1302;  
 RN [1]  
 RP SEQUENCE FROM N.A.



RC STRAIN-V288;  
 RA Bolken T.C., Franke C.A., Hruby D.E., Zeller G.O.;  
 RT "Identification of New Intergenic and Intragenic Integration Sites for  
 RT Foreign Gene Expression in Recombinant *Streptococcus gordonii*  
 RT Strains.";  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF210773; AAG39000.1; -  
 DR InterPro: IPR003501; PTS\_1IB.  
 DR Pfam: PF02302; PTS\_1IB; 1.  
 KW Transferase.  
 FT NON\_TER  
 SQ SEQUENCE 251 AA; 27174 MW; B85CCDFDFFD8D33 CRC64;

Query Match  
 Best Local Similarity 100.0%; Score 33; DB 2; Length 251;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGCTSG 6  
 Db 157 GGCTSG 162

RESULT 5  
 O9NZT0 PRELIMINARY; PRT; 256 AA.  
 AC O9NZT0;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE PLUNC.  
 GN PLUNC.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bingle C.D.;  
 RT "Human PLUNC, a novel upper airway enriched protein."  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF172993; AAF70860.1; -  
 SQ SEQUENCE 256 AA; 26712 MW; EDF152PBD7EC5E5C CRC64;

Query Match  
 Best Local Similarity 100.0%; Score 33; DB 4; Length 256;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGCTSG 6  
 Db 80 GGCTSG 85

RESULT 6  
 O9NP55 PRELIMINARY; PRT; 256 AA.  
 AC O9NP55;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE LUNG PROTEIN (NASOPHARYNGEAL CARCINOMA-RELATED PROTEIN) (TRACHEAL  
 DE EPITHELIUM ENRICHED PROTEIN) (BAA9G10.5) (LUNG PROTEIN (LOC51297)).  
 GN LUNG OR PLUNC OR BAA9G10.5.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA TISSUE=LUNG;  
 RC Watanabe T.;  
 RA "Novel gene specifically expressed in lung.";

RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA He Z.W., Yao K.T., Xu L.G., Ren C.P., Lan K., Xie L., Zhang L.,  
 RA Liu W.D., Zhou W., Wang L.;  
 RT "Cloning a new gene related to nasopharyngeal carcinoma."  
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20472055; PubMed=11018263;  
 RA Bingle C.D., Bingle L.;  
 RT "Characterization of the human plunc gene, a gene product with an  
 RT upper airways and nasopharyngeal restricted expression pattern."  
 RL Biochim. Biophys. Acta 1493:363-367(2000).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Tracey A.;  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB024937; BAA93633.1; -  
 DR EMBL: AF158745; AAF82622.1; -  
 DR EMBL: AF214562; AAG13653.1; -  
 DR EMBL: AL121901; CAC03549.1; -  
 SQ SEQUENCE 256 AA; 26712 MW; EDF152PBC35315BC CRC64;

Query Match  
 Best Local Similarity 100.0%; Score 33; DB 4; Length 256;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGCTSG 6  
 Db 80 GGCTSG 85

RESULT 7  
 O9VYL0 PRELIMINARY; PRT; 275 AA.  
 AC O9VYL0;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE CG11085 PROTEIN.  
 GN CG11085.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew K.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
 RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Hoch S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gong F., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heilmann T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Jiang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Mikhlov G., Mishina N.V., Mobarry C., Morris J., Moshneff A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
RA Palazzolo M., Pittman K.A., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Sinden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RT "The genome sequence of *Drosophila melanogaster*."  
RL Science 287:2185-2195(2000).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
CC -1- SIMILARITY: TO OTHER HOMEOBOX DOMAINS.  
DR EMBL: AF003489; AAF48182.1; -.  
DR HSSP: P22808; INK3.  
DR FLYBase: FBgn0030408; CG11085.  
DR InterPro: IPR002952; Eggshell.  
DR InterPro: IPR001356; Homeobox.  
DR Pfam: PF00046; homeobox; 1.  
DR PRINTS: PR01328; EGGSHLL.  
DR PRINTS: PR00024; HOMEOBOX.  
DR SMART: SM00389; HOX; 1.  
DR PROSITE: PS00027; HOMEOBOX\_1; 1.  
DR PROSITE: PS50071; HOMEOBOX\_2; 1.  
KW DNA-binding; Homeobox; Nuclear protein.  
KW DNA-binding; Homeobox; Nuclear protein.  
SQ SEQUENCE 275 AA; 28185 MW; AB3150295C656EE3 CRC64;

Query Match 100.0%; Score 33; DB 5; Length 275;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GCGTSG 6  
DB 119 GCGTSG 124

RESULT 8  
062416 PRELIMINARY: PRT; 279 AA.  
AC 062416:  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-JAN-1999 (TREMBLrel. 09, Last annotation update)  
DE Y2P5A.5 PROTEIN.  
GN Y2P5A.5.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Gardner A.;  
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berts M., Coulson A.,  
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,  
RA Smailson N., Smith A., Sonhammer E., Staden R., Sulston J.,  
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
RA Watson A., Weinstein L., Wilkinson-Sproat J., Wohlman P.;

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*."  
RL Nature 368:32-38(1994).  
DR EMBL: AL021479; CA116324.1; -.  
SQ SEQUENCE 279 AA; 30291 MW; 1A51F6351F08E4E CRC64;

Query Match 100.0%; Score 33; DB 5; Length 279;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GCGTSG 6  
DB 206 GCGTSG 211

RESULT 9  
080450 PRELIMINARY: PRT; 289 AA.  
ID 080450:  
AC 080450:  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE PUTATIVE GTL1 PROTEIN.  
GN F16M14.18.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA Rounsley S.D., Kaul S., Lin X., Ketchum K.A., Crosby M.L.,  
RA Brannon R.C., Sykes S.M., Mason T.M., Kerlavage A.R., Adams M.D.,  
RA Somerville C.R., Venter J.C.;  
RT "Arabidopsis thaliana chromosome II BAC F16M14 genomic sequence."  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AC003028; AAC27174.1; -.  
DR InterPro: IPR001005; MYB\_DNA\_bind.  
DR PROSITE: PS50090; MYB\_3; 1.  
DR SMART: SM00395; SANT; 1.  
SQ SEQUENCE 289 AA; 34307 MW; CEE46E6022724754 CRC64;

Query Match 100.0%; Score 33; DB 10; Length 289;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GCGTSG 6  
DB 142 GCGTSG 147

RESULT 10  
09LYP4 PRELIMINARY: PRT; 303 AA.  
ID 09LYP4:  
AC 09LYP4:  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE HYPOHETICAL 34.2 KDA PROTEIN.  
GN T28J14.160.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Bevan M., Murphy G., Ridley P., Hudson S., Bancroft I., Meves H.W.,  
RA Rudd S., Lemcke K., Mayer K.F.X.;  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.

[2]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AL163652; CAB87278.1; -  
DR InterPro: IPR003103; BAG.  
DR InterPro: IPR003130; GED.  
DR InterPro: IPR000626; Ubiquitin.  
DR Pfam: PF02179; BAG; 1.  
DR SMART: SM00264; BAG; 1.  
DR SMART: SM00302; GED; 1.  
DR PROSITE: PS50053; UBQUITIN\_2; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 303 AA; 34244 MW; EB3464BD401A9D37 CRC64;

Query Match 100.0%; Score 33; DB 10; Length 303;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
Db 14 GCGTSG 19  
IIIIII

RESULT 11  
ID O9Y1H0 PRELIMINARY; PRT; 315 AA.  
AC O9Y1H0;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)  
DE REVERSE TRANSCRIPTASE-LIKE PROTEIN (FRAGMENT).  
OS Amblyomma americanum.  
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;  
OC Parasitiformes; Ixodida; Ixodidae; Amblyomma.  
OX NCBI\_TaxId=6943;  
RN (1)  
RP SEQUENCE FROM N.A.  
RA Bunlik J., Hall L., Barbour A.G.;  
RT "Reverse transcriptase-like sequence in the tick Amblyomma americanum";  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE TRANSCRIPTASE).  
CC EMBL: AF134586; AAD39461.1; -  
DR InterPro: IPR000477; RVTse.  
DR Pfam: PF00078; rvt; 1.  
KW RNA-directed DNA polymerase.  
FT NON\_TER 1 315  
FT 1 315  
SQ SEQUENCE 315 AA; 35459 MW; A101E22ABF33FE14 CRC64;

Query Match 100.0%; Score 33; DB 5; Length 315;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
Db 175 GCGTSG 180  
IIIIII

RESULT 12  
ID 075953 PRELIMINARY; PRT; 348 AA.  
AC 075953;  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)  
DE HEAT SHOCK PROTEIN HSP40-3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxId=9606;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC TISSUE-COLON;  
RA Chen M.S., Laszlo A., Roti Roti J.L.;  
RT "Differential response of members in the hsp40 family to various stress agents in mammalian cells";  
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF088982; AAC35860.1; -  
DR HSP; P25685; 1HDJ.  
DR InterPro: IPR002939; DnaJ\_C.  
DR InterPro: IPR001623; DnaJ\_N.  
DR Pfam: PF00226; DnaJ; 1.  
DR Pfam: PF01556; DnaJ\_C; 1.  
DR SMART: SM00271; DnaJ; 1.  
DR PROSITE: PS00636; DnaJ\_1; 1.  
DR PROSITE: PS50076; DnaJ\_2; 1.  
KW Heat shock.  
SQ SEQUENCE 348 AA; 39133 MW; DC9FE45DE4FDCFC CRC64;

Query Match 100.0%; Score 33; DB 4; Length 348;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
Db 75 GCGTSG 80  
IIIIII

RESULT 13  
ID O9GE43 PRELIMINARY; PRT; 427 AA.  
AC O9GE43;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)  
DE RIBULOSE BIPHOSPHATE CARBOXYLASE LARGE CHAIN (EC 4.1.1.39) (RUBISCO LARGE SUBUNIT) (FRAGMENT).  
GN RBCL.  
OS Pseudoscleropodium purum.  
OG Chloroplast.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;  
OC Bryopsida; Bryidae; Hypnales; Brachytheciaceae; Pseudoscleropodium.  
OX NCBI\_TaxId=94508;  
RN (1)  
RP SEQUENCE FROM N.A.  
RA De Luna E., Buck W.R., Akiyama H., Arikawa T., Tsubota H., Gonzalez D., Newton A.E., Shaw A.J., Gonzalez D.;  
RT "Ordinal Phylogeny within the Hypnophyalean Pleurocarpous Mosses Inferred from Cladistic Analyses of Three Chloroplast DNA Sequence Data Sets: trnL-F, rps4, and rbcL";  
RL Bryologist 103:242-256(2000).  
CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME ACTIVE SITE (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) -> 2 3-PHOSPHO-D-GLYCERATE.  
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) -> 3-PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.  
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST (BY SIMILARITY).  
DR EMBL: AF233567; AAG27527.1; -  
DR InterPro: IPR000685; Rubisco\_large.  
DR Pfam: PF00016; Rubisco\_large; 1.  
DR PROSITE: PS00157; RUBISCO\_LARGE; 1.  
KW Carbon dioxide fixation; Chloroplast; lyase; Monooxygenase; Oxidoreductase; Photorespiration; Photosynthesis.

FT NON\_TER 1 1  
FT NON\_TER 427 427  
SO SEQUENCE 427 AA: 47365 MW: 7F837679470879E CRC64:

Query Match 100.0%; Score 33; DB 8; Length 427;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
DB 383 GCGTSG 387

RESULT 14  
O9TMM3 PRELIMINARY; PRT; 431 AA.

AC O9TMM3;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE RIBULOSE BIPHOSPHATE CARBOXYLASE LARGE CHAIN (EC 4.1.1.39) (RUBISCO LARGE SUBUNIT) (FRAGMENT).  
GN RBCL.  
OS Selaginella denticulata.  
OC Chloroplast.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Lycopodiophyta; Isoetopsida; Selaginellales; Selaginellaceae;  
OC Selaginella.  
OX NCBI\_TaxID=50508;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Korrall P., Therrien J.P., Kenrick P.;  
RT "Phylogeny of Selaginellaceae: evaluation of generic/subgeneric relationships based on rbcL gene sequences.";  
RL Int. J. Plant Sci. 160:585-594(1999).  
CC -1- FUNCTION: RUBISCO CATALYSES TWO REACTIONS: THE CARBOXYLATION OF D-RIBULOSE 1,5-BIPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME ACTIVE SITE (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BIPHOSPHATE + CO(2) -> 2 3-PHOSPHO-D-GLYCERATE.  
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BIPHOSPHATE + O(2) -> 3-PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.  
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.  
DR EMBL; AJ010853; CAB54079.1; -.  
DR HSSP; P00880; RBUL.  
DR InterPro; IPR000685; RUBISCO\_large.  
DR Pfam; PF00016; RUBISCO\_large.  
DR PROSITE; PS00157; RUBISCO\_large; 1.  
KW Carbon dioxide fixation; Chloroplast; Lyase; Monooxygenase; Oxidoreductase; Photorespiration; Photosynthesis.  
FT NON\_TER 1 1  
FT NON\_TER 431 431  
SO SEQUENCE 431 AA: 46651 MW: EB671D290AB6CFE5 CMC64:

Query Match 100.0%; Score 33; DB 8; Length 431;  
Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
DB 382 GCGTSG 387

RESULT 15  
O03112 PRELIMINARY; PRT; 449 AA.

OY 1

AC O03112;  
DT 01-JUL-1997 (TREMBLrel. 04, Created)  
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE RIBULOSE BIPHOSPHATE CARBOXYLASE LARGE CHAIN (EC 4.1.1.39) (RUBISCO LARGE SUBUNIT) (FRAGMENT).  
GN RBCL.  
OS Lophocolea heterophylla.  
OC Chloroplast.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta; Jungermanniales; Geocalycinae; Lophocoleaceae; Lophocolea.

OY 1 GCGTSG 6  
DB 392 GCGTSG 397

RESULT 16  
O31681 PRELIMINARY; PRT; 451 AA.

AC O31681;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE RIBULOSE BIPHOSPHATE CARBOXYLASE LARGE CHAIN (EC 4.1.1.39) (RUBISCO LARGE SUBUNIT) (FRAGMENT).  
GN RBCL.  
OS Andreea rupestris.  
OC Chloroplast.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta; Andreeopsida; Andreeales; Andreeaceae; Andreeae.  
OX NCBI\_TaxID=13797;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Manhart J.R.;  
RT "Phylogenetic analysis of green plant rbcL sequences.";  
Mol. Phylogenet. Evol. 3:114-127(1994).

Query Match 100.0%; Score 33; DB 8; Length 449;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
DB 392 GCGTSG 397

RESULT 16  
O31681 PRELIMINARY; PRT; 451 AA.

AC O31681;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE RIBULOSE BIPHOSPHATE CARBOXYLASE LARGE CHAIN (EC 4.1.1.39) (RUBISCO LARGE SUBUNIT) (FRAGMENT).  
GN RBCL.  
OS Andreea rupestris.  
OC Chloroplast.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta; Andreeopsida; Andreeales; Andreeaceae; Andreeae.  
OX NCBI\_TaxID=13797;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Manhart J.R.;  
RT "Phylogenetic analysis of green plant rbcL sequences.";  
Mol. Phylogenet. Evol. 3:114-127(1994).

CC -1- FUNCTION: RUBISCO CATALYSES TWO REACTIONS: THE CARBOXYLATION OF D-  
CC RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC  
CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF  
CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH  
CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME  
CC ACTIVE SITE.  
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) -> 2 3-  
CC PHOSPHO-D-GLYCERATE.  
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) -> 3-  
CC PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.  
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.  
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.  
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.  
DR EMBL: L13473; AAA53403.1; -  
DR Mendel: 3036; Andru:rbcl;3036.  
DR InterPro: IPR000685; Rubisco\_large.  
DR InterPro: IPR001412; rRNA-synt\_1.  
DR Pfam: PF00016; Rubisco\_large; 1.  
DR PROSITE: PS00178; AA\_TRNA\_LIGASE\_1; UNKNOWN\_1.  
KM Carbon dioxide fixation: Chloroplast; Lyase; Monooxygenase;  
KM Oxidoreductase; Photorespiration; Photosynthesis.  
FT ACT\_SITE 192 192 BINDING OF CO(2) ACTIVATES THE ENZYME.  
FT ACT\_SITE 193 193 BINDING OF CO(2) ACTIVATES THE ENZYME.  
FT NON\_TER 451 451  
SQ SEQUENCE 451 AA; 49815 MW; 65987158F1C7B6C5 CRC64;  
  
QY 1 GCGTSG 6  
Db 394 GCGTSG 399  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
RESULT 17  
Q9BCK9 PRELIMINARY; PRT; 468 AA.  
AC Q9BCK9;  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE RIBULOSE-1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE LARGE SUBUNIT  
DE (FRAGMENT).  
GN RBCL.  
OS Desmarestia sp. Seto Inland Seas.  
OG Chloroplast.  
OC Eukaryota; stramenopiles; Phaeophyceae; Desmarestiales;  
OC Desmarestiaceae; Desmarestia.  
OX NCBI\_TaxID=114257;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SETO INLAND SEAS;  
RA Kawai H., Sasaki H.;  
RT "Molecular phylogeny of brown algal genera Akkesiphycus and Halosiphon  
RT (Lammariales), resulting in the circumscription of the new families  
RT Akkesiphycaceae and Halosiphonaceae.";  
RL Phycologia 39:416-428(2000).  
DR EMBL: AB037141; BAB32897.1; -  
KW Chloroplast.  
FT NON\_TER 1 1  
SQ SEQUENCE 468 AA; 51314 MW; 61A1F7C2DFA82D45 CRC64;  
  
Query Match 100.0%; Score 33; DB 8; Length 451;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GCGTSG 6  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
NCBI\_TaxID=9606;

Db 386 GCGTSG 391  
RESULT 18  
Q9GHW8 PRELIMINARY; PRT; 469 AA.  
ID Q9GHW8;  
AC Q9GHW8;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE RIBULOSE BISPHOSPHATE CARBOXYLASE LARGE CHAIN (EC 4.1.1.39) (RUBISCO  
DE LARGE SUBUNIT) (FRAGMENT).  
GN RBCL.  
OS Gomphandra javanica.  
OG Chloroplast.  
CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
CC Asteridae; euasterids II; Icacinaceae; Gomphandra.  
OX NCBI\_TaxID=124965;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Savolainen V., Fay M.F., Albach D.C., Backlund A., van der Bank M.,  
RA Cameron K.M., Johnson S.A., Lledo M.D., Pintaud J.C., Powell M.,  
RA Sheahan M.C., Soltis D.E., Soltis P.S., Weston P., Whitten W.M.,  
RA Wurdack K.J., Chase M.W.;  
RT "Phylogeny of the eudicots: a nearly complete familial analysis based  
RT on rbcL gene sequences.";  
RL Kew Bull. 55:257-309(2000).  
CC -1- FUNCTION: RUBISCO CATALYSES TWO REACTIONS: THE CARBOXYLATION OF D-  
CC RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC  
CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF  
CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH  
CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME  
CC ACTIVE SITE (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) -> 2 3-  
CC PHOSPHO-D-GLYCERATE.  
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) -> 3-  
CC PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.  
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.  
DR EMBL: AJ402954; CAC04312.1; -  
DR InterPro: IPR000685; Rubisco\_large.  
DR Pfam: PF00016; Rubisco\_large; 1.  
DR PROSITE: PS00157; RUBISCO\_LARGE; 1.  
KM Carbon dioxide fixation; Chloroplast; Lyase; Monooxygenase;  
KM Oxidoreductase; Photorespiration; Photosynthesis.  
FT NON\_TER 1 1  
FT NON\_TER 469 469  
SQ SEQUENCE 469 AA; 51914 MW; BC18693698AD95FE CRC64;  
  
QY 1 GCGTSG 6  
Db 396 GCGTSG 401  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
RESULT 19  
Q9HLJ0 PRELIMINARY; PRT; 470 AA.  
ID Q9HLJ0;  
AC Q9HLJ0;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE HUPF3B.  
OS Homo sapiens (Human).  
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC Mammalia; Euteria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxID=9606;

RN (1)  
 RE SEQUENCE FROM N.A.  
 RX PUBMED-11163187;  
 RA Lykke-Andersen J., Shu M.-D., Steitz J.A.;  
 RT "Human Upf proteins target an mRNA for nonsense-mediated decay when  
 bound downstream of a termination codon."  
 RL Cell 103:1121-1131(2000)  
 DR EMBL: AY013251; AAC48511.1;  
 SQ SEQUENCE 470 AA; 56213 MW; C012044D4509957F CRC64;

Query Match 100.0%; Score 33; DB 4; Length 470;  
 Best Local Similarity 100.0%; Pred. No. 3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
 |||||  
 DB 26 GCGTSG 31

RESULT 20  
 O9TM74 PRELIMINARY; PRT; 475 AA.  
 ID O9TM74;  
 AC O9TM74;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE RIBULOSE BISPHOSPHATE CARBOXYLASE LARGE CHAIN (EC 4.1.1.39) (RUBISCO  
 LARGE SUBUNIT).  
 GN RBCL.  
 OS Rhynchosostegium pallidifolium.  
 OC Chloroplast.  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;  
 OC Bryopsida; Bryidae; Hypnales; Brachytheciaceae; Rhynchosostegium.  
 RX NCBI\_TaxID=90343;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-HIGASHI-HIROSHIMA;  
 RA Tsubota H., Nakao N., Arikawa T., Yamaguchi T., Higuchi M.,  
 RT "A preliminary phylogeny of Hypnales (Musci) as inferred from  
 chloroplast rbcL sequence data."  
 RL Bryol. Res. 7:233-248(1999).  
 CC -1- FUNCTION: RUBISCO CATALYSES TWO REACTIONS: THE CARBOXYLATION OF D-  
 RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC  
 CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF  
 THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH  
 REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME  
 ACTIVE SITE (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) -> 2 3-  
 PHOSPHO-D-GLYCERATE.  
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) -> 3-  
 PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.  
 CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST (BY SIMILARITY).  
 DR EMBL: AB024944; BAA86213.1;  
 DR InterPro: IPR000685; RUBISCO\_large.  
 DR Pfam: PF00016; RUBISCO\_large.1.  
 DR PROSITE: PS00157; RUBISCO\_LARGE; 1.  
 KW Oxidoreductase; Photoregulation; Photosynthesis;  
 SO SEQUENCE 475 AA; 52740 MW; 4D7BA2D5C815B7E7 CRC64;

Query Match 100.0%; Score 33; DB 8; Length 475;  
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
 |||||  
 DB 403 GCGTSG 408

RESULT 21  
 O9TM63 PRELIMINARY; PRT; 475 AA.  
 ID O9TM63;  
 AC O9TM63;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE RIBULOSE BISPHOSPHATE CARBOXYLASE LARGE CHAIN (EC 4.1.1.39) (RUBISCO  
 LARGE SUBUNIT).  
 GN RBCL.  
 OS Platyhypnidium riparioides.  
 OC Chloroplast.  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;  
 OC Bryopsida; Bryidae; Hypnales; Brachytheciaceae; Platyhypnidium.  
 RX NCBI\_TaxID=98942;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H. TSUBOTA 2210;  
 RA Tsubota H., Nakao N., Arikawa T., Yamaguchi T., Higuchi M.,  
 RT "A preliminary phylogeny of Hypnales (Musci) as inferred from  
 chloroplast rbcL sequence data."  
 RL Bryol. Res. 7:233-248(1999).  
 CC -1- FUNCTION: RUBISCO CATALYSES TWO REACTIONS: THE CARBOXYLATION OF D-  
 RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC  
 CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF  
 THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH  
 REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME  
 ACTIVE SITE (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) -> 2 3-  
 PHOSPHO-D-GLYCERATE.  
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) -> 3-  
 PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.  
 CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST (BY SIMILARITY).  
 DR EMBL: AB029385; BAA86206.1;  
 DR InterPro: IPR000685; RUBISCO\_large.  
 DR Pfam: PF00016; RUBISCO\_large.1.  
 DR PROSITE: PS00157; RUBISCO\_LARGE; 1.  
 KW Carbon dioxide fixation; Chloroplast; Lyase; Monooxygenase;  
 SO SEQUENCE 475 AA; 52740 MW; C3F551355A19ECC CRC64;

Query Match 100.0%; Score 33; DB 8; Length 475;  
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
 |||||  
 DB 403 GCGTSG 408

RESULT 22  
 O9B217 PRELIMINARY; PRT; 483 AA.  
 ID O9B217;  
 AC O9B217;  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE UPF3X.  
 GN UPF3X.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RX NCBI\_TaxID=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=20565755; PUBMED=11113196;  
 RA Serin G., Gersappe A., Black J.D., Aronoff R., Maguati L.E.;  
 RT "Identification and characterization of human orthologues to

RT saccharomyces cerevisiae upf2 protein and upf3 protein (Caenorhabditis  
RT elegans smg-4).  
RL Mol. Cell. Biol. 21:209-223(2001).  
DR EMBL: AF318576; AAG60691.1; -  
SQ SEQUENCE 483 AA; 57762 MW; F5A8A395783D1A69 CRC64;

Query Match 100.0%; Score 33; DB 4; Length 483;  
Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGGTSG 6  
|||||  
DB 26 GGGTSG 31

RESULT 23  
Q9VSE9 PRELIMINARY; PRT; 484 AA.

AC Q9VSE9;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE CG7404 PROTEIN.  
GN CG7404.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN-BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoshino R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang J., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazer J.C., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Bokorova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
RA Butris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Dudin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jaitli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.W., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paley J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Pui Y., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spler E., Spredler A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weisscock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.; -  
RT \*The genome sequence of Drosophila melanogaster.\*;  
RL Science 287:2185-2195(2000).  
CC -I- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).

CC -I- SIMILARITY: TO C4-TYPE STEROID RECEPTOR ZINC FINGER FAMILY.  
DR EMBL: AE003556; AAF50473.1; -  
DR HSSP: P03372; IHCP.  
DR FlyBase: FBgn0035849; CG7404.  
DR InterPro: IPR000536; Hormone\_rec\_1lg.  
DR InterPro: IPR001723; Steroidhormone\_receptor.  
DR InterPro: IPR001628; zf-C4.  
DR Pfam: PF00104; hormone\_rec.1.  
DR Pfam: PF00105; zf-C4.1.  
DR PRINTS: PRO0398; STRDHORMONER.  
DR PRINTS: PRO0047; STROIDFINGER.  
DR SMART: SM00430; HOL1.1.  
DR SMART: SM00399; znf\_C4.1.  
DR PROSITE: PS00031; NUCLEAR\_RECEPTOR.1.  
KW DNA-binding; Nuclear protein; Receptor; Transcription regulation;  
KW Zinc-finger.  
SQ SEQUENCE 484 AA; 53245 MW; 252E3AD07C8A311 CRC64;

Query Match 100.0%; Score 33; DB 5; Length 484;  
Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGGTSG 6  
|||||  
DB 87 GGGTSG 92

RESULT 24  
Q9A4M2 PRELIMINARY; PRT; 503 AA.

AC Q9A4M2;  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE TRYPTOPHAN HALOGENASE, PUTATIVE.  
GN CC2808.  
OS Caulobacter crescentus.  
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
OC Caulobacter.  
OX NCBI\_TaxID=69394;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21173698; PubMed=11259647;  
RA Nieman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,  
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwin M.L., Haft D.H.,  
RA Kolonay J.F., Smit J., Craven M.B., Khouli H., Shetty J., Berry K.,  
RA Ullrich T., Tran K., Wolf A., Vamathevan J., Ermlaeva M., White O.,  
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.; -  
RT \*Complete genome sequence of Caulobacter crescentus.\*;  
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
DR EMBL: AE005946; AAK24772.1; -  
DR TIGR: CC2808; -  
KW Complete proteome.  
SQ SEQUENCE 503 AA; 56671 MW; 15BD6A3BAF8FEFE2 CRC64;

Query Match 100.0%; Score 33; DB 2; Length 503;  
Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGGTSG 6  
|||||  
DB 14 GGGTSG 19

RESULT 25  
Q9AA21 PRELIMINARY; PRT; 504 AA.

AC Q9AA21;  
DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE TRYPOTOPHAN HALOGENASE, PUTATIVE.  
GN CC0787.  
OS Caulobacter crescentus.  
OC Bacteria: Proteobacteria; alpha subdivision; Caulobacter group;  
CC Caulobacter  
NCBI\_Taxid=69394;  
RN  
[1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21173698; Pubmed=11259647;  
RA Nielsen W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,  
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
RA Deboy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,  
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,  
RA Ullrich J., Tran K., Wolf A., Yamashita J., Ermolaeva M., White O.,  
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
RT \*Complete genome sequence of Caulobacter crescentus.\*;  
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
DR EMBL; AF005754; AAK22772.1;  
KW Complete proteome.  
SQ SEQUENCE 504 AA; 55610 MW; 342C629BD65DA1C4 CRC64;  
  
Query Match 100.0%; Score 33; DB 2; Length 504;  
Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GCGTSG 6  
DB 13 GCGTSG 18  
  
RESULT 26  
027403 PRELIMINARY; PRT; 504 AA.  
AC 027403;  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE GCM PROTEIN.  
GN GCM OR GUIDE/GCM OR CG12245.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
NCBI\_Taxid=7227;  
RN  
[1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96016097; Pubmed=7553844;  
RA Hosoya T., Takizawa K., Nitta K., Hotta Y.;  
RT \*Glial cells missing: a binary switch between neuronal and glial  
RT determination in Drosophila.\*;  
RL Cell 82:1025-1036(1995).  
RN  
[2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96016096; Pubmed=7553843;  
RA Jones B.W., Fetter R.D., Tear G., Goodman C.S.;  
RT \*Glial cells missing: a genetic switch that controls glial versus  
RT neuronal fate.\*;  
RL Cell 82:1013-1023(1995).  
RN  
[3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98022869; Pubmed=9356176;  
RA Bernardoni R., Vivancos V., Giangrande A.;  
RT \*glide/gcm is expressed and required in the scavenger cell lineage.\*;  
RN Dev. Biol. 191:118-130(1997).  
RN  
[4]  
RP SEQUENCE FROM N.A.  
RA Verdeli A., Giangrande A.;  
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.

DR EMBL; D64040; BAA10905.1; -;  
DR EMBL; U34039; AAC46912.1; -;  
DR EMBL; U81164; AAC47808.1; -;  
DR TRANSFAC; T02302; -;  
DR FlyBase; FBgn0014179; gcm.  
DR InterPro; IPR003902; GCM\_moltlf.  
SQ SEQUENCE 504 AA; 56168 MW; 07D7DC22C4944C4F CRC64;  
  
Query Match 100.0%; Score 33; DB 5; Length 504;  
Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GCGTSG 6  
DB 430 GCGTSG 435  
  
RESULT 27  
09VLAS PRELIMINARY; PRT; 504 AA.  
AC 09VLAS;  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE GCM PROTEIN.  
GN GCM OR CG12245.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
NCBI\_Taxid=7227;  
RN  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY.  
RX MEDLINE=20196006; Pubmed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Morten J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Adair J.F., Agbayani A., An H.-J., Andrews-Plamkoc C., Baldwin D.,  
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borokova D., Borchan M.R., Bouck J., Broksrein P., Brothier P.,  
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Delcher A., Deng Z., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster A., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeigwe C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., Meperson D.,  
RA Merkulov G., Milshina N.V., Mobery C., Morris J., Moshell A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spter E., Spradling A.C., Stapleton M., Strong M., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT \*The genome sequence of Drosophila melanogaster.\*;



RL Science 287:2185-2195(2000).  
 DR EMBL: AE003625; AAF52790.1; -  
 DR FlyBase: FBgn0014179; gcm.  
 DR InterPro: IPR003902; GCM\_motif.  
 SO SEQUENCE 504 AA; 56202 MW; 0ED74C2D3B9BCEB0 CRC64;

Query Match 100.0%; Score 33; DB 5; Length 504;  
 Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGGTSG 6  
 |||||  
 DB 430 GGGTSG 435

RESULT 28  
 O9W0N5 PRELIMINARY; PRT; 528 AA.  
 AC O9W0N5;  
 DT 01-MAY-2000 (TReMBLrel. 13, Created)  
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
 DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)  
 DE CG13894 PROTEIN.  
 GN CG13894.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 ON NCBI\_TaxID=7227;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers J.H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abri J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,  
 RA Ballieu R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Bortova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup I.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasro P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paolel J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reiert K., Remington K., Saunders R.D.C., Schneider F., Shen H.,  
 RA Shue B.C., Sidenkiamos I., Stimpson M., Skupski M.P., Smith T.,  
 RA Spiteras R., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhou G., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003469; AAF47408.1; -

DR FlyBase: FBgn0035157; CG13894.  
 SO SEQUENCE 528 AA; 59427 MW; 911EEF7AECF1C450 CRC64;

Query Match 100.0%; Score 33; DB 5; Length 528;  
 Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGGTSG 6  
 |||||  
 DB 411 GGGTSG 416

RESULT 29  
 O82435 PRELIMINARY; PRT; 559 AA.  
 AC O82435;  
 DT 01-NOV-1998 (TReMBLrel. 08, Created)  
 DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE (R) (+) MANDELONITRILE LYASE ISOFORM MDL5 PRECURSOR.  
 GN MDL5.  
 OS Prunus serotina (Black cherry).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucotsids I; Rosales; Rosaceae; Prunus.  
 ON NCBI\_TaxID=23207;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Hu Z., Poulton J.E.;  
 RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AF053886; AAC61982.1; -  
 DR Mendel: 33024; Pruse: 2903; 33024.  
 DR InterPro: IPR00172; GNC\_oxred.  
 DR Pfam: PF00732; GNC\_oxred. 1.  
 DR PROSITE: PS00623; GNC\_OXRED\_1; 1.  
 DR PROSITE: PS00624; GNC\_OXRED\_2; 1.  
 KW Lyase.  
 SO SEQUENCE 559 AA; 61090 MW; 2368FEB75359C1C5 CRC64;

Query Match 100.0%; Score 33; DB 10; Length 559;  
 Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGGTSG 6  
 |||||  
 DB 60 GGGTSG 65.

RESULT 30  
 O9F97 PRELIMINARY; PRT; 563 AA.  
 AC O9F97;  
 DT 01-MAR-2001 (TReMBLrel. 16, Created)  
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE MANDELONITRILE LYASE-LIKE PROTEIN.  
 GN Arabidopsis thaliana (Mouse-ear cress).  
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucotsids II; Brassicales; Brassicaceae; Arabidopsids.  
 ON NCBI\_TaxID=3702;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=COLUMBIA;  
 RX MEDLINE=99087489; PubMed=9872454;  
 RA Nakamura Y., Sato S., Asamizu E., Kaneko T., Kotani H., Miyajima N.,  
 RA Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. VII.  
 RT Sequence features of the regions of 1,013,767 bp covered by sixteen  
 RT DNA Res. 5:297-308(1998).  
 RL

DR EMBL: AB015478; BAB1043.1; -  
 DR InterPro: IPR000172; GMC\_oxred.  
 DR Pfam: PF00732; GMC\_oxred; 1.  
 DR PROSITE: PS00624; GMC\_OXRED\_2; 1.  
 KW Lyase.  
 SO SEQUENCE 563 AA; 61979 MW; 4803987007889FE9 CRC64;

Query Match 100.0%; Score 33; DB 10; Length 563;  
 Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
 |||||  
 DB 52 GCGTSG 57

## RESULT 31

O99Y17 PRELIMINARY; PRT; 565 AA.

AC O99Y17;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DE PUNITIVE PTS SYSTEM, LACTOSE-SPECIFIC COMPONENT IIBC (EC 2.7.1.69).  
 GN LACE OR SPY1917.  
 OS Streptococcus pyogenes.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1314;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SF370:  
 RA MEDLINE-21192684, PubMed-11296296.  
 RA Ferretti J.J., Moshan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,  
 RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,  
 RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,  
 RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;  
 RT \*Complete genome sequence of an M1 strain of Streptococcus pyogenes.\*;  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).  
 DR EMBL: AE006616; AAK34621.1; -  
 KW Transferase; Complete proteome.  
 SO SEQUENCE 565 AA; 61209 MW; 122FA0043FDA2D84 CRC64;

Query Match 100.0%; Score 33; DB 2; Length 565;  
 Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
 |||||  
 DB 471 GCGTSG 476

## RESULT 32

O99S77 PRELIMINARY; PRT; 570 AA.

AC O99S77;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DE PTS SYSTEM, LACTOSE-SPECIFIC IIBC COMPONENT.  
 GN LACE OR SAI1992.  
 OS Streptococcus aureus subsp. aureus N315.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Staphylococcus.  
 OX NCBI\_TaxID=158879;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
 RA Cui L., Oguchi A., Aoki K.I., Nagai Y., Iino J., Ito T., Kanamori M.,  
 RA Matsumaru H., Matsuyama A., Murakami H., Hosoyama A., Mizutani-Ul Y.,  
 RA Takahashi N.K., Sawano T., Inoue R.I., Kaito C., Sekimizu K.,

RA Hirakawa H., Kuhara S., Goto S., Yabuzaki J., Kanehisa M.,  
 RA Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T., Hattori M.,  
 RA Ogasawara N., Hayashi H., Hiramatsu K.;  
 RT \*Whole genome sequencing of methicillin-resistant Staphylococcus  
 aureus.\*;  
 RL Lancet 357:1225-1240(2001).  
 DR EMBL: AP003136; BAB43282.1; -  
 KW Complete proteome.  
 SO SEQUENCE 570 AA; 62414 MW; 11B6FCC1B716028E CRC64;

Query Match 100.0%; Score 33; DB 2; Length 570;  
 Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
 |||||  
 DB 476 GCGTSG 481

## RESULT 33

O82784 PRELIMINARY; PRT; 574 AA.

AC O82784;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DE MD14 PROTEIN.  
 GN MD14.  
 OS Prunus serotina (Black cherry).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids I; Rosales; Rosaceae; Prunus.  
 OX NCBI\_TaxID=23207;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Hu Z., Poulton J.E.;  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Hu Z., Poulton J.E.;  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF053885; AAC61981.1; -  
 DR EMBL: AF053884; AAC61980.1; -  
 DR EMBL: AF043187; AAD02265.1; -  
 DR EMBL: AF043186; AAD02265.1; -  
 DR Mendel: 33350; Pruse; 2903; 33350.  
 DR Mendel: 35927; Pruse; 2903; 35927.  
 DR Mendel: 35928; Pruse; 2903; 35928.  
 DR InterPro: IPR00172; GMC\_oxred.  
 DR Pfam: PF00732; GMC\_oxred; 1.  
 DR PROSITE: PS00624; GMC\_OXRED\_2; 1.  
 KW Lyase.  
 SO SEQUENCE 574 AA; 61682 MW; BCDFAE1DC5D2C539 CRC64;

Query Match 100.0%; Score 33; DB 10; Length 574;  
 Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
 |||||  
 DB 61 GCGTSG 66

## RESULT 34

P94865 PRELIMINARY; PRT; 577 AA.

AC P94865;  
 DT 01-MAY-1997 (TREMBLrel. 03, Created)  
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)  
 DE LACE.  
 SO SEQUENCE 574 AA; 61682 MW; BCDFAE1DC5D2C539 CRC64;

GN LACE.  
 OS Lactobacillus casei.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 GN Lactobacillus.  
 OC NCBI\_TaxID=1582;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 393 (PL215-);  
 RX MEDLINE=97218691; PubMed=9066115;  
 RA Gosalbes M., Monedero V., Albert C., Perez-Martinez G.;  
 RT "Establishing a model to study the regulation of the lactose operon in  
 RT Lactobacillus casei.";  
 RL FEBS Microbiol. Lett. 148:83-89(1997).  
 DR EMBL; Z80834; CAB02556.1; -  
 DR InterPro: IPR003352; PTS\_EIIC.  
 DR InterPro: IPR003501; PTS\_IIB.  
 DR Pfam: PF02302; PTS\_IIB.1  
 DR Pfam: PF02378; PTS\_EIIC.1.  
 SQ SEQUENCE 577 AA; 62367 MW; 32DB7B35E55607CB CRC64;

Query Match 100.0%; Score 33; DB 2; Length 577;  
 Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
 |||||  
 DB 485 GCGTSG 490

RESULT 35  
 ID Q9P6V0 PRELIMINARY; PRT; 578 AA.  
 AC Q9P6V0;  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE ORIGIN RECOGNITION COMPLEX SUBUNIT 2 RELATED PROTEIN.  
 GN 15E6.30.  
 OS Neurospora crassa.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Sordariales; Sordariaceae; Neurospora.  
 OX NCBI\_TaxID=5141;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Schulte U., Algin V., Hoheisel J., Brandt P., Fartmann B., Holland R.,  
 RA Nyakatura G., Mewes H.W., Mannhaupt G.;  
 RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA German Neurospora genome project;  
 RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL353822; CAB88634.1; -  
 DR InterPro: IPR001990; Granin.  
 DR PROSITE: PS00422; GRANIN\_1; UNKNOWN\_1.  
 SQ SEQUENCE 578 AA; 63706 MW; E7010DA0176F0FC4 CRC64;

Query Match 100.0%; Score 33; DB 3; Length 578;  
 Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
 |||||  
 DB 521 GCGTSG 526

RESULT 36  
 ID Q9SINI PRELIMINARY; PRT; 618 AA.  
 AC Q9SINI;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)

DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE AT2G42580 PROTEIN.  
 GN AT2G42580.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. COLUMBIA;  
 RX MEDLINE=20083487; PubMed=10617197;  
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,  
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,  
 RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,  
 RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,  
 RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,  
 RA Salzberg S.L., Fraser C.M., Venter J.C.;  
 RT "Sequence and analysis of chromosome II of Arabidopsis thaliana.";  
 RL Nature 402:761-768(1999).  
 DR EMBL; AC007087; AAD22995.1; -  
 DR InterPro: IPR001440; TPR.  
 DR Pfam: PF00515; TPR.6  
 DR SMART: SM00028; TPR.4.  
 SQ SEQUENCE 618 AA; 66897 MW; C0E1A18CA4212D86 CRC64;

Query Match 100.0%; Score 33; DB 10; Length 618;  
 Best Local Similarity 100.0%; Pred. No. 4.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
 |||||  
 DB 203 GCGTSG 208

RESULT 37  
 ID 056765 PRELIMINARY; PRT; 668 AA.  
 AC 056765;  
 DT 01-JUN-1998 (Tremblrel. 06, Created)  
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE UL52. PROTEIN.  
 GN UL52.  
 OS Human cytomegalovirus (strain Towne).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Betaherpesvirinae; Cytomegalovirus.  
 OX NCBI\_TaxID=10363;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-TOWNE;  
 RX MEDLINE=98241707; PubMed=9573236;  
 RA Krosky P.M., Underwood M.R., Turk S.R., Feng K.W.-H., Jain R.K.,  
 RA Ptak R.G., Westerman A.C., Biron K.K., Townsend L.B., Drach J.C.;  
 RT "Resistance of human cytomegalovirus to benzimidazole ribonucleosides  
 RT maps to two open reading frames: UL89 and UL56.";  
 RL J. Virol. 72:4721-4728(1998).  
 DR EMBL; AF047521; AAC40816.1; -  
 DR InterPro: IPR002597; Herpes\_env.  
 DR Pfam: PF01673; Herpes\_env.1.  
 SQ SEQUENCE 668 AA; 74154 MW; 049848D7A52596CC CRC64;

Query Match 100.0%; Score 33; DB 12; Length 668;  
 Best Local Similarity 100.0%; Pred. No. 4.5e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
 |||||  
 DB 422 GCGTSG 427

```

RESULT 38
O9ASR6 PRELIMINARY: PRT; 691 AA.
AC O9ASR6:
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ATG42580/F14N22.15.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae:
OC eucotsids II: Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_Taxid=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Meyers M.C., Shinn P., Banh J.,
RA Bowser L., Carninci P., Chung M.K., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jones T., Kamlya A., Karlin-Neumann G., Kawai J., Lam B.,
RA Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M., Nguyen M.,
RA Palm C.J., Pham P.K., Quach H.L., Sakano H., Sakurai T., Satou M.,
RA Seki M., Southwick A., Toriumi M., Yamada K., Yu G., Shinozaki K.,
RA Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis cDNA clones."
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF367321; AAK32908.1;
DR EMBL: AF367321; AAK32908.1;
SQ SEQUENCE 691 AA; 75461 MW; ABF8F609DDB6E19 CRC64;

```

```

Query Match 100.0%; Score 33; DB 10; Length 691;
Best Local Similarity 100.0%; Pred. No. 4.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 GCGTSG 6
|111111
DB 203 GCGTSG 208

```

```

RESULT 39
O9DBY4 PRELIMINARY: PRT; 809 AA.
AC O9DBY4:
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE 1200009022RIK PROTEIN.
CN 1200009022RIK.
OS Mus musculus (Mouse).
OC Eukaryota: Metazoa: Chordata: Vertebrata: Euteleostomi:
OC Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=LUNG;
RA MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Iizawa M., Nishii K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fieischmann W., Gaasterland T., Glass C., King B., Kochiya H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staahl F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Ronald M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamlya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Kombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodiguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;

```

```

RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL: AK004681; BAB23469.1;
DR MGD: MGI:1914123; 1200009022RIK.
DR InterPro: IPR001611; LRR.
DR InterPro: IPR000483; LRR_Cterm.
DR InterPro: IPR000372; LRR_Nterm.
DR InterPro: IPR003592; LRR_out.
DR InterPro: IPR003591; LRR_cyp.
DR Pfam: PF00560; LRR_10.
DR Pfam: PF01463; LRRCT; 1.
DR PRINTS: PRO0019; LEURICHRPT.
DR SMART: SM00370; LRR; 8.
DR SMART: SM00082; LRRCT; 1.
DR SMART: SM00013; LRRNT; 1.
DR SMART: SM00369; LRR_TIP; 11.
SQ SEQUENCE 809 AA; 88809 MW; 5E86F55BA8AE419FA CRC64;

```

```

Query Match 100.0%; Score 33; DB 11; Length 809;
Best Local Similarity 100.0%; Pred. No. 5.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 GCGTSG 6
|111111
DB 790 GCGTSG 795

```

```

RESULT 40
O9BUL1 PRELIMINARY: PRT; 817 AA.
AC O9BUL1:
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NUCLEAR RECEPTOR NHR-48.
OS Caenorhabditis elegans.
OC Eukaryota: Metazoa: Nematoda: Chromadorea: Rhabditida: Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Bogan A., Maiba C.V., Yamamoto K., Cohen F., Sluder A.E.;
RT "Caenorhabditis elegans nuclear receptor sequences exhibit biophysical
RT compatibility with the ligand-binding domain fold."
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF332205; AAK17976.1;
KM Receptor.
SQ SEQUENCE 817 AA; 90706 MW; 036211C94A913BB9 CRC64;

```

```

Query Match 100.0%; Score 33; DB 5; Length 817;
Best Local Similarity 100.0%; Pred. No. 5.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 GCGTSG 6
|111111
DB 64 GCGTSG 69

```

```

Search completed: February 4, 2002, 08:02:21
Job time: 125 sec

```



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 4, 2002, 08:03:03 ; Search time 38.23 Seconds  
(Without alignments)  
11.625 Million cell updates/sec

Title: us-09-642-660-11  
Perfect score: 30  
Sequence: 1 GSLGGS 6

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Database : Listing first 100 summaries

1: /SID88/gcgdata/geneseq/AA1980.DAT.\*  
2: /SID88/gcgdata/geneseq/AA1981.DAT.\*  
3: /SID88/gcgdata/geneseq/AA1982.DAT.\*  
4: /SID88/gcgdata/geneseq/AA1983.DAT.\*  
5: /SID88/gcgdata/geneseq/AA1984.DAT.\*  
6: /SID88/gcgdata/geneseq/AA1985.DAT.\*  
7: /SID88/gcgdata/geneseq/AA1986.DAT.\*  
8: /SID88/gcgdata/geneseq/AA1987.DAT.\*  
9: /SID88/gcgdata/geneseq/AA1988.DAT.\*  
10: /SID88/gcgdata/geneseq/AA1989.DAT.\*  
11: /SID88/gcgdata/geneseq/AA1990.DAT.\*  
12: /SID88/gcgdata/geneseq/AA1991.DAT.\*  
13: /SID88/gcgdata/geneseq/AA1992.DAT.\*  
14: /SID88/gcgdata/geneseq/AA1993.DAT.\*  
15: /SID88/gcgdata/geneseq/AA1994.DAT.\*  
16: /SID88/gcgdata/geneseq/AA1995.DAT.\*  
17: /SID88/gcgdata/geneseq/AA1996.DAT.\*  
18: /SID88/gcgdata/geneseq/AA1997.DAT.\*  
19: /SID88/gcgdata/geneseq/AA1998.DAT.\*  
20: /SID88/gcgdata/geneseq/AA1999.DAT.\*  
21: /SID88/gcgdata/geneseq/AA2000.DAT.\*  
22: /SID88/gcgdata/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30	100.0	156	22	AA000035
2	30	100.0	277	21	AAV82699
3	30	100.0	281	21	AAV90256
4	30	100.0	287	22	AA000097
5	30	100.0	411	21	AA10987
6	30	100.0	411	22	AA071099
7	30	100.0	416	21	AA011003
8	30	100.0	416	22	AA071115
9	30	100.0	416	21	AA071157
10	30	100.0	433	21	AA033306
11	30	100.0	484	21	AA033305

12	30	100.0	931	17	AA04867	Transferrin bindin
13	30	100.0	1346	18	AA031346	Rat tumour suppres
14	30	100.0	1356	18	AA031347	Rat tumour suppres
15	30	93.3	1596	15	AA057055	HTLV peptide. HTLV
16	30	93.3	97	21	AA001209	Human secreted pro
17	30	93.3	119	22	AA074477	Human colon cancer
18	30	93.3	633	19	AA075774	Amino acid sequenc
19	30	93.3	633	19	AA075775	Amino acid sequenc
20	30	93.3	1892	21	AA050497	Arabidopsis thaila
21	30	93.3	1940	21	AA050495	Arabidopsis thaila
22	30	93.3	1944	21	AA050496	Arabidopsis thaila
23	30	90.0	67	20	AA074755	Mycobacterium spec
24	30	90.0	91	21	AA027140	Zea mays protein f
25	30	90.0	121	21	AA045378	Human secreted pro
26	30	90.0	126	22	AA059979	Human endometrium
27	30	90.0	126	22	AA059504	Human protein sequ
28	30	90.0	163	19	AA053009	Mus musculus I-mfc
29	30	90.0	173	20	AA035432	Chlamydia pneumoni
30	30	90.0	188	21	AA024867	Arabidopsis thaila
31	30	90.0	190	21	AA039461	Human secreted pro
32	30	90.0	192	21	AA014530	Arabidopsis thaila
33	30	90.0	196	19	AA081735	Arabidopsis thaila
34	30	90.0	196	19	AA064370	M. tuberculosis im
35	30	90.0	196	20	AA039167	M. tuberculosis im
36	30	90.0	196	20	AA039167	M. tuberculosis im
37	30	90.0	233	21	AA056851	M. tuberculosis an
38	30	90.0	235	20	AA06887	M. tuberculosis re
39	30	90.0	246	19	AA053007	Human prostate can
40	30	90.0	251	19	AA053008	Human mature recep
41	30	90.0	254	22	AA029789	Mus musculus I-mfc
42	30	90.0	290	01	AA079285	Mus musculus I-mfb
43	30	90.0	308	22	AA012221	Consensus eukaryot
44	30	90.0	318	18	AA044200	Rat hem oxygenase
45	30	90.0	318	18	AA037554	Human PK0403 poly
46	30	90.0	322	21	AA014529	Human mature recep
47	30	90.0	326	22	AA066886	Human RAGE polyep
48	30	90.0	326	22	AA029788	Arabidopsis thaila
49	30	90.0	332	21	AA052130	HHH4720 polyep
50	30	90.0	336	21	AA014528	Human serine/threo
51	30	90.0	336	21	AA024866	Human Receptor to
52	30	90.0	337	22	AA029235	Arabidopsis thaila
53	30	90.0	340	18	AA044199	Arabidopsis thaila
54	30	90.0	342	21	AA033753	Human protein sequ
55	30	90.0	342	21	AA024865	Human soluble rece
56	30	90.0	349	18	AA035849	Human RAGE polyep
57	30	90.0	349	20	AA035433	Arabidopsis thaila
58	30	90.0	354	20	AA035434	Human CDS for use
59	30	90.0	368	21	AA018343	Chlamydia pneumoni
60	30	90.0	378	21	AA059149	Chlamydia pneumoni
61	30	90.0	394	22	AA083099	Plasmodium falcipa
62	30	90.0	404	22	AA081925	Human GAK protein
63	30	90.0	446	22	AA065005	S. epidermidis ope
64	30	90.0	449	22	AA094059	Extracellular cfr
65	30	90.0	493	22	AA024056	Human secreted pro
66	30	90.0	494	21	AA041839	Human protein sequ
67	30	90.0	501	22	AA067811	Human Est encoded
68	30	90.0	501	22	AA067811	Human ORX ORF1603
69	30	90.0	510	22	AA067165	Amino acid sequenc
70	30	90.0	510	22	AA065080	Arabidopsis histon
71	30	90.0	513	12	AA014304	Gene #13 associate
72	30	90.0	513	15	AA048065	Insect steroid rec
73	30	90.0	513	18	AA018876	D. melanogaster ul
74	30	90.0	513	19	AA031320	Insect XR2C recept
75	30	90.0	584	12	AA014670	D. melanogaster XR
76	30	90.0	584	12	AA014671	Truncated poly Ig-
77	30	90.0	629	22	AA027121	Truncated poly Ig-
78	30	90.0	654	22	AA068405	C-terminally trunc
79	30	90.0	656	22	AA068407	Amino acid sequenc
80	30	90.0	657	22	AA068407	C glutamic prote
81	30	90.0	673	21	AA096449	Amino acid sequenc
82	30	90.0	708	22	AA075616	Forward transcrip
83	30	90.0	773	17	AA031177	Human colon cancer
84	30	90.0	900	22	AA030338	Rabbit poly-immuno
					AA030338	A multifunctional
					AA030338	Saccharomyces cere

85	27	90.0	900	22	AA820177	S. cerevisiae mult
86	27	90.0	900	22	AA820178	S. cerevisiae mult
87	27	90.0	900	22	AA820179	S. cerevisiae mult
88	27	90.0	967	21	AA844536	Virulence gene pro
89	27	90.0	985	20	AA889419	Moraxella catarrha
90	27	90.0	985	20	AA889415	Moraxella catarrha
91	27	90.0	1000	20	AA889414	Moraxella catarrha
92	27	90.0	1000	20	AA889418	Moraxella catarrha
93	27	90.0	1305	20	AA879274	Protein kinase GAK
94	27	90.0	1311	20	AA879275	Protein kinase GAK
95	27	90.0	1342	12	AA813833	HER-3 epithelial g
96	26	86.7	9	21	AA898526	WT1 derived immuno
97	26	86.7	9	21	AA898741	WT1 derived immuno
98	26	86.7	9	22	AA861858	Human WT1 immunoge
99	26	86.7	9	22	AA862073	Mouse WT1 immunoge
100	26	86.7	10	22	AA895959	Human complementat

## ALIGNMENTS

## RESULT 1

ID AAM00035 standard; Protein: 156 AA.

AC AAM00035;

DT 28-SEP-2001 (first entry)

DE Endoxylglucan transferase sequence #103.

KW Moss; carbohydrate metabolism related protein; CMPR; sugar; cofactor;

KM fine chemical production; carbohydrate; polysaccharide.

OS Physcomitrella patens.

PN WO200144476-A2.

PD 21-JUN-2001.

PF 14-DEC-2000; 2000WO-EP12697.

PR 16-DEC-1999; 99US-0171101.

PA (BADI ) BASF PLANT SCI GMBH.

PI Lerchl J, Renz A, Ehrhardt T, Reindl A, Cirpus P, Bischoff F;

PI Frank M, Freund A, Duwenig E, Schmidt R, Reski R;

DR WPI: 2001-398155/42.

DR N-PSDB; AAH88721.

XX Novel moss nucleic acid molecules encoding a carbohydrate metabolism  
 PT related protein useful for modulating production of fine chemicals such  
 PT as carbohydrates, cofactors and enzymes from microorganisms and plants

PS Claim 30; Page 115; 133pp; English.

XX This invention relates to nucleic acid molecules AAH88708 - AAH88796  
 CC isolated from Physcomitrella patens (a moss), which encode carbohydrate  
 CC metabolism related proteins (CMPR) represented in AAM00022 - AAM00110.  
 CC Included in the invention is a vector containing the CMPR cDNA, and a  
 CC host cell transformed with the vector. The host cell (a microorganism,  
 CC Corynebacterium or Brevibacterium, moss or algae or a plant cell) is  
 CC useful for producing a fine chemical such as carbohydrates, cofactors  
 CC and/or enzymes. The nucleic acid molecules are suitable for modifying a  
 CC carbohydrate production system in a host, e.g., microorganisms and  
 CC plants. They are also useful to identify those DNA sequences and enzymes  
 CC in other species which are useful to modify the biosynthesis of starch,  
 CC cell wall polysaccharides and sugars. The nucleic acid molecules may be  
 CC utilised in the genetic engineering of Corynebacterium glutamicum and the  
 CC related Brevibacterium species and Acetobacter xylinum and Chlorella to

CC make it a better or more efficient producer of one or more fine  
 CC chemicals. Mutagenesis of one or more CMPRs may also result in CMPRs  
 CC having altered activities which indirectly impact the production of one  
 CC or more desired fine chemicals from plants. Primers AAH88705 - AAH88707  
 CC are used in the sequencing of the CMPR cDNA sequences of the invention.

SQ Sequence 156 AA;

Query Match 100.0%; Score 30; DB 22; Length 156;

Best Local Similarity 100.0%; Pred. No. 1.3e+02; Indels 0; Gaps 0;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSIGGS 6

DB 2 GSIGGS 7

## RESULT 2

ID AAY82699 standard; Protein: 277 AA.

AC AAY82699;

DT 10-AUG-2000 (first entry)

DE Tick salivary gland antigen related immunogen protein sequence.

KW Tick; vaccine; infection; salivary gland antigen; immunogen;

KM serine protease; cysteine protease; blood sucker.

OS Haemaphysalis longicornis.

FH Key Location/Qualifiers

FT Misc-difference 269 /note="encoded by T"

FN JP2000083677-A.

PD 28-MAR-2000.

PF 17-SEP-1998; 98JP-0281932.

PR 17-SEP-1998; 98JP-0281932.

PA (FARB ) BAYER KK.

PI WPI: 2000-296340/26.

DR N-PSDB; AAA29614.

PT A gene encoding tick salivary gland antigen - useful as a vaccine for  
 PT protecting animals from tick-carried infections

PS Claim 2; Page 7; 29pp; Japanese.

XX The present sequence represents a tick salivary gland antigen related  
 CC immunogen. The present invention also describes a tick derived serine  
 CC protease and a tick derived cysteine protease. A nucleotide sequence  
 CC encoding any of the above proteins can be used in a vaccine against  
 CC tick carried infections for domestic animals such as cattle.

SQ Sequence 277 AA;

Query Match 100.0%; Score 30; DB 21; Length 277;

Best Local Similarity 100.0%; Pred. No. 2.4e+02; Indels 0; Gaps 0;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSIGGS 6

DB 39 GSIGGS 44



RESULT 3  
AA90256  
ID AAY90256 standard. Protein: 281 AA.  
XX  
XX  
AC AAY90256;  
XX  
DT 19-SEP-2000 (first entry)  
XX  
DE Streptococcus equi fibronectin binding protein.  
XX  
XX Fibronectin binding protein; SFS; vaccine; horse; strangles; therapy;  
XX equine upper respiratory tract disease; S. equi infection.  
XX  
OS Streptococcus equi.  
XX  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 23..30 /note="encoded by GTGGGA"  
FT Misc-difference 46..54 /note="encoded by TCCACGAC"  
FT Misc-difference 70..71 /note="encoded by TCACACCGA AAAAGATTA CTGACGCTT  
FT /note="AAGCTTCAG AT"  
FT Misc-difference 75..83 /note="encoded by AATCGCCG"  
FT Misc-difference 98..106 /note="encoded by CTCACCAA"  
FT Misc-difference 122..123 /note="encoded by CAAGGCTGA GAGAGTTA CAGCTTGT  
FT /note="TCAGTCAATC AC"  
FT Misc-difference 127..134 /note="encoded by CAGAT"  
FT Misc-difference 150..158 /note="encoded by GCGAAAAA"  
FT Misc-difference 174..175 /note="encoded by GGTTCAGGT TGAATGTTA TAATCAAGC  
FT /note="CGTCTTGC TT"  
FT Misc-difference 179..187 /note="encoded by AATGCACAG"  
FT Misc-difference 202..210 /note="encoded by TTAGATCTT"  
FT Misc-difference 226..227 /note="encoded by GGTAAATGATG AATCACCATA ACTTAAAGAC  
FT /note="GAACACCCCT AC"  
FT Misc-difference 233..239 /note="encoded by CGCTAT"  
FT Misc-difference 235..262 /note="encoded by CGAAG"  
FT Misc-difference 278..279 /note="encoded by GCGGATAGAG GTGTAAAGAG AGGATACGAC  
TT TATCTTACA AT"  
XX  
XX WO200037496-A1.  
XX  
XX 29-JUN-2000.  
XX  
XX 21-DEC-1999; 99WO-SE02448.  
XX  
XX 22-DEC-1998; 98SE-0004491.  
XX  
XX (GUS/) GUS B.  
XX (LIND/) LINDMARK H.  
XX (JACO/) JACOBSSON K.  
XX (FRYK/) FRYKBERG L.  
XX  
XX Guss B, Lindmark H, Jacobsson K, Frykberg L;  
XX  
XX WPI: 2000-442641/38.  
XX  
XX N-PSDB; AAA30869.  
XX  
XX  
XX New protein useful for preparation of vaccines for treatment of  
XX strangles caused by Streptococcus equi infection, is able to bind to  
XX mammalian fibronectin.

XX  
PS Claim 2; Page 29-30; 34pp; English.  
XX  
XX This sequence represents the Streptococcus equi fibronectin binding  
CC protein of the invention, designated SFS. SFS binds specifically to  
CC mammalian fibronectin or its analogues or fragments. The protein, its  
CC analogues or fragments may be used for the preparation of a vaccine that  
CC protects horses against strangles (a world-wide distributed and serious  
CC disease of the equine upper respiratory tract) caused by S. equi  
CC infection. The antibody and/or antiserum may also be used for the  
CC prophylactic or therapeutic treatment of S. equi infection in mammal,  
CC especially horses. The use of vaccines containing the fibronectin binding  
CC protein provides a more effective protection against S. equi infections,  
CC with fewer side effects.  
XX  
XX  
SQ Sequence 281 AA.  
XX  
XX  
XX Query Match 100.0%; Score 30; DB 21; Length 281;  
XX Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
XX Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX 1 GSUGGS 6  
XX 111111  
XX 40 gslggs 45  
XX  
RESULT 4  
AA00097  
ID AA00097 standard. Protein: 287 AA.  
XX  
XX AA00097;  
XX  
XX 28-SEP-2001 (first entry)  
XX  
XX Endoxylglucan transferase sequence #165.  
XX  
XX  
XX Moss; carbohydrate metabolism related protein; CMP; sugar; cofactor;  
XX fine chemical production; carbohydrate; polysaccharide.  
XX  
XX Physcomitrella patens.  
XX  
XX WO200144476-A2.  
XX  
XX 21-JUN-2001.  
XX  
XX 14-DEC-2000; 2000WO-EP12697.  
XX  
XX 16-DEC-1999; 99US-0171101.  
XX  
XX (BAD1) BASF PLANT SCI GMBH.  
XX  
XX Lerchl J, Renz A, Ehrhardt T, Reindl A, Cirpus P, Bischoff F;  
XX Frank M, Freund A, Duwenig E, Schmidt R, Reski R;  
XX  
XX WPI: 2001-398155/42.  
XX  
XX N-PSDB; AAH88783.  
XX  
XX Novel moss nucleic acid molecules encoding a carbohydrate metabolism  
XX related protein useful for modulating production of fine chemicals such  
XX as carbohydrates, cofactors and enzymes from microorganisms and plants  
XX  
XX  
XX Claim 30; Page 127-128; 133pp; English.  
XX  
XX This invention relates to nucleic acid molecules AAH88708 - AAH88796  
XX isolated from Physcomitrella patens (a moss), which encode carbohydrate  
XX metabolism related proteins (CMP) represented in AA00097 - AA00110.  
XX Included in the invention is a vector containing the CMP cDNA, and a  
XX host cell transformed with the vector. The host cell (a microorganism,  
XX Corynebacterium or Brevibacterium, moss or algae or a plant cell) is  
XX useful for producing a fine chemical such as carbohydrates, cofactors  
XX and/or enzymes. The nucleic acid molecules are suitable for modifying a

CC carbohydrate production system in a host, e.g., microorganisms and  
CC plants. They are also useful to identify those DNA sequences and enzymes  
CC in other species which are useful to modify the biosynthesis of starch,  
CC cell wall polysaccharides and sugars. The nucleic acid molecules may be  
CC utilized in the genetic engineering of *Corynebacterium glutamicum* and the  
CC related *Brevibacterium* species and *Acetobacter xylinum* and *Chlorella* to  
CC make it a better or more efficient producer of one or more fine  
CC chemicals. Mutagenesis of one or more CMRs may also result in CMRs  
CC having altered activities which indirectly impact the production of one  
CC or more desired fine chemicals from plants. Primers AAH88705 - AAH80707  
CC are used in the sequencing of the CMR cDNA sequences of the invention.  
XX

SO Sequence 287 AA;

Query Match 100.0%; Score 30; DB 22; Length 287;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSIGCS 6  
|11111  
DB 2 gslggs 7

RESULT 5  
AAB10987  
ID AAB10987 standard; Protein: 411 AA.  
XX  
AC AAB10987;  
XX  
DT 16-FEB-2001 (first entry)  
XX  
DE M. crenulata hemocyanin KLN2 domain c.  
XX  
KW Hemocyanin; cytosolic; virucide; antibacterial; antiparasitic;  
KW immunomodulatory; antihypertensive; gene therapy; tumor; treatment;  
KW infection; schistosomiasis; carcinoma; bladder; epithelium; ovary;  
KW breast; bronchi; colon-rectum; hypertension; vaccine; cocaine misuse;  
KW pharmaceutical carrier.  
XX  
OS Megathura crenulata.  
XX  
PN WO200055192-A2.  
XX  
PD 21-SEP-2000.  
XX  
PE 17-MAR-2000; 2000WO-EP02410.  
XX  
PR 17-MAR-1999; 99DE-1011971.  
PR 20-AUG-1999; 99DE-1039578.  
XX  
PA (BIOS-) BIOSYN ARZNEIMITTEL GMBH.  
XX  
PI Markl J, Altenhein B, Lleb B, Stiefel T;  
XX  
DR WPI; 2000-587517/55.  
XX  
PT New nucleic acid encoding hemocyanin, useful for gene therapy of tumors  
XX and for recombinant production of fusion proteins for vaccination  
XX  
PS Claim 21; Page 116-117; 163pp; German.  
XX

CC This invention describes a novel nucleic acid (I) containing a sequence  
CC that encodes hemocyanin (II), a domain of (I) or its fragment with the  
CC immunological properties of at least one domain of (II). The products of  
CC the invention have cytostatic, virucide, antibacterial, antiparasitic,  
CC immunomodulatory and antihypertensive activity. (II), and constructs  
CC additionally containing antigen-encoding sequences, are useful in gene  
CC therapy of tumors. Polypeptides encoded by (I) are useful for treating  
CC parasitic or viral infections and tumors, particularly schistosomiasis  
CC and carcinoma (of bladder, epithelium, ovary, breast, bronchi or  
CC colon-rectum), also hypertension, as vaccines, for treating cocaine  
CC misuse and very generally as carriers for pharmaceuticals, e.g.

CC cytostatics. They may also be used to generate antibodies (Ab). Probes  
CC based on (I) and Ab are useful for detecting tumor-specific DNA in a cell  
CC (by detecting specific binding to cellular DNA or proteins), particularly  
CC where associated with the types of carcinoma listed above. Hemocyanin  
CC can be produced recombinantly, relatively inexpensively and in adequate  
CC amounts, eliminating the need to culture gastropods. When used as a  
CC carrier, (II) significantly increases the half-life of the attached  
CC pharmaceutical, by inhibiting ultrafiltration in the kidneys.  
XX

SO Sequence 411 AA;

Query Match 100.0%; Score 30; DB 21; Length 411;  
Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSIGCS 6  
|11111  
DB 390 gslggs 395

RESULT 6  
AAB71099  
ID AAB71099 standard; Protein: 411 AA.  
XX  
AC AAB71099;  
XX  
DT 24-AUG-2001 (first entry)  
XX  
DE M. crenulata KLN2 domain c.  
XX  
KW HeH1; KLN1; hemocyanin; cytosolic; vasotropic; vaccine; gene therapy;  
KW tumor; parasite infection; viral infection; antitumor agent;  
KW abnormal blood pressure; bladder carcinoma; epithelial carcinoma;  
KW ovarian carcinoma; mammary carcinoma; bronchial carcinoma;  
KW colon carcinoma; cocaine addiction.  
XX  
OS Megathura crenulata.  
XX  
PN WO200114536-A2.  
XX  
PD 01-MAR-2001.  
XX  
PE 21-AUG-2000; 2000WO-EP08129.  
XX  
PR 20-AUG-1999; 99DE-1039578.  
XX  
PA (BIOS-) BIOSYN ARZNEIMITTEL GMBH.  
XX  
PI Markl J, Altenhein B, Lleb B, Stiefel T;  
XX  
DR WPI; 2001-191646/19.  
XX  
PT New nucleic acid encoding hemocyanin, useful for gene therapy of tumors  
XX and for recombinant production of fusion proteins for vaccination  
XX  
PS Claim 25; Page 133-134; 180pp; German.  
XX

CC This invention describes a novel nucleic acid molecule (NI) encoding  
CC hemocyanin, a hemocyanin domain, or a functional fragment of it with  
CC the immunological properties of at least one domain of hemocyanin, and  
CC which comprises at least one intron is new. The products of the invention  
CC have cytostatic and vasotropic activity and can be used in vaccines or  
CC for gene therapy. Compositions comprising nucleic acids of the invention  
CC are useful in treating tumors. Compositions comprising hemocyanin  
CC polypeptide are useful in treating parasite and viral infections and as  
CC an antitumor agent. It is also useful in treating abnormal blood  
CC pressure, bladder carcinomas, epithelial carcinomas, ovarian carcinomas,  
CC mammary carcinomas, bronchial carcinomas, and colon carcinomas.  
CC It may also be used to cure cocaine addiction.  
XX

SO Sequence 411 AA;

Query Match 100.0%; Score 30; DB 22; Length 411;  
 Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSLSGS 6  
 |||||  
 Db 390 gsl9gs 395

## RESULT 7

AAB11003  
 ID AAB11003 standard; Protein: 416 AA.

AC AAB11003;

DT 16-FEB-2001 (first entry)

DE M. crenulata hemocyanin KLM2 domain C'.

XX Hemocyanin; cytosolic; virucide; antibacterial; antiparasitic;

KW immunomodulatory; antihypertensive; gene therapy; tumor; treatment;

KM infection; schistosomiasis; carcinoma; bladder; epithelium; ovary;

KW breast; bronchi; colon-rectum; hypertension; vaccine; cocaine misuse;

XX pharmaceutical carrier.

OS Megathura crenulata.

PN WO20005192-A2.

PD 21-SEP-2000.

PF 17-MAR-2000; 2000MO-EP02410.

XX 17-MAR-1999; 99DE-1011971.

PR 20-AUG-1999; 99DE-1039578.

XX (BIOS-) BIOSYN ARZNEIMITTEL GMBH.

PA Markl J, Altenhein B, Lieb B, Stiefel T;

PI MPI; 2000-587517/55.

DR New nucleic acid encoding hemocyanin, useful for gene therapy of tumors

PT and for recombinant production of fusion proteins for vaccination -

XX Claim 21; Page 142-143; 163pp; German.

XX This invention describes a novel nucleic acid (I) containing a sequence

CC that encodes hemocyanin (II), a domain of (I) or its fragment with the

CC immunological properties of at least one domain of (II). The products of

CC the invention have cytosolic, virucide, antibacterial, antiparasitic,

CC immunomodulatory and antihypertensive activity. (I), and constructs

CC additionally containing antigen-encoding sequences, are useful in gene

CC therapy of tumors. Polypeptides encoded by (I) are useful for treating

CC parasitic or viral infections and tumors, particularly schistosomiasis

CC and carcinoma (of bladder, epithelium, ovary, breast, bronchi or

CC colon-rectum), also hypertension, as vaccines, for treating cocaine

CC misuse and very generally as carriers for pharmaceuticals, e.g.

CC cytosolics. They may also be used to generate antibodies (Ab). Probes

CC based on (I) and Ab are useful for detecting tumor-specific DNA in a cell

CC (by detecting specific binding to cellular DNA or proteins), particularly

CC where associated with the types of carcinoma listed above. Hemocyanins

CC can be produced recombinantly, relatively inexpensively and in adequate

CC amounts, eliminating the need to culture gastropods. When used as a

CC carrier, (II) significantly increases the half-life of the attached

CC pharmaceutical, by inhibiting ultrafiltration in the kidneys.

CC Sequence 416 AA;

Query Match 100.0%; Score 30; DB 21; Length 416;

Best Local Similarity 100.0%; Pred. No. 3.7e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSLSGS 6  
 |||||  
 Db 398 gsl9gs 403

## RESULT 8

AAB71115  
 ID AAB71115 standard; Protein: 416 AA.

AC AAB71115;

DT 24-AUG-2001 (first entry)

DE M. crenulata KLM2 domain C'.

XX HTH1; KLM1; hemocyanin; cytosolic; vasotropic; vaccine; gene therapy;

KW tumor; parasite infection; viral infection; antitumor agent;

KW abnormal blood pressure; bladder carcinoma; epithelial carcinoma;

KW ovarian carcinoma; mammary carcinoma; bronchial carcinoma;

XX colon carcinoma; cocaine addiction.

OS Megathura crenulata.

PN WO200114536-A2.

PD 01-MAR-2001.

PF 21-AUG-2000; 2000MO-EP08129.

XX 20-AUG-1999; 99DE-1039578.

XX (BIOS-) BIOSYN ARZNEIMITTEL GMBH.

PA Markl J, Altenhein B, Lieb B, Stiefel T;

PI MPI; 2001-191646/19.

DR New nucleic acid encoding hemocyanin, useful for gene therapy of tumors

PT and for recombinant production of fusion proteins for vaccination -

XX Claim 25; Page 159-160; 180pp; German.

XX This invention describes a novel nucleic acid molecule (N1) encoding

CC hemocyanin, a hemocyanin domain, or a functional fragment of it with

CC the immunological properties of at least one domain of hemocyanin, and

CC which comprises at least one intron is new. The products of the invention

CC have cytosolic and vasotropic activity and can be used in vaccines or

CC for gene therapy. Compositions comprising nucleic acids of the invention

CC are useful in treating tumors. Compositions comprising hemocyanin

CC polypeptide are useful in treating parasite and viral infections and as

CC an antitumor agent. It is also useful in treating abnormal blood

CC pressure, bladder carcinomas, epithelial carcinomas, ovarian carcinomas,

CC mammary carcinomas, bronchial carcinomas, and colon carcinomas. It may

CC also be used to cure cocaine addiction.

XX Sequence 416 AA;

Query Match 100.0%; Score 30; DB 22; Length 416;

Best Local Similarity 100.0%; Pred. No. 3.7e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSLSGS 6  
 |||||  
 Db 398 gsl9gs 403

## RESULT 9

AAB71157  
 ID AAB71157 standard; Protein: 416 AA.

XX



PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142035.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0145977.  
PR 13-JUL-1999; 99US-0145442.  
PR 14-JUL-1999; 99US-0145624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144352.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 20-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 21-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145813.  
PR 27-JUL-1999; 99US-0145918.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149929.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.

PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158028.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 18-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160981.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 100.0%; Score 30; DB 21; Length 433;  
Best Local Similarity 100.0%; Pred. NO. 3; Be+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSLGGS 6  
|||||  
Db 2 gslggs 7

RESULT 11  
AAG33305  
ID AAG33305 standard; Protein; 484 AA.  
XX  
AC AAG33305;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Zea mays protein fragment SEQ ID NO: 40335.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence; corn.  
XX  
XX Zea mays subsp. mays.  
XX  
XX EPI033405-A2.  
XX  
XX 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX

PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123348.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 04-MAY-1999; 99US-0132048.  
PR 05-MAY-1999; 99US-0132407.  
PR 06-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 07-MAY-1999; 99US-0132486.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134376.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 14-JUN-1999; 99US-0138847.  
PR 16-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140389.  
PR 23-JUN-1999; 99US-0140384.  
PR 24-JUN-1999; 99US-0140354.  
PR 28-JUN-1999; 99US-0140665.  
PR 29-JUN-1999; 99US-0140823.  
PR 30-JUN-1999; 99US-0140991.  
PR 01-JUL-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142055.  
PR 08-JUL-1999; 99US-0142390.  
PR 09-JUL-1999; 99US-0142803.  
PR 12-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 20-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147028.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147316.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148341.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 16-AUG-1999; 99US-0148684.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153070.  
PR 15-SEP-1999; 99US-0153758.  
PR 16-SEP-1999; 99US-0154039.  
PR 16-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.

PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158368.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 100.0%; Score 30; DB 21; Length 484;  
 Best Local Similarity 100.0%; Pred. No. 4.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLGGS 6  
 |||||  
 Db 53 gslggs 58

RESULT 12  
 AAM04867  
 ID AAM04867 standard; Protein; 931 AA.

AC AAM04867;  
 XX  
 XX 19-JAN-1997 (first entry)  
 DT  
 XX Transferrin binding protein 1 gene.  
 DE  
 XX Transferrin binding protein 1; Tbp1; vaccine; antibody; diagnosis;  
 KW swine pleuropneumonia.  
 KW  
 OS Actinobacillus pleuropneumoniae strain 1371 serotype 1.  
 XX  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..22  
 FT /label= sig-peptide  
 XX  
 XX EP733708-A2.  
 PN  
 XX 25-SEP-1996.  
 PD  
 XX 21-MAR-1996; 96EP-0870033.  
 PF  
 XX 24-MAR-1995; 95ES-0000592.  
 PR  
 XX (HPR-) LAB HIPRA SA.  
 PA  
 XX

PI Daban M, Espuna E, Medrano A, Querol E;  
 XX WPI; 1996-427056/43.  
 DR N-PSDS; AAT38071.  
 XX  
 XX Actinobacillus pleuropneumoniae transferrin binding protein 1 - for  
 PT production of antibodies useful diagnostically and in universal  
 FT vaccine against porcine pleuropneumonia  
 XX  
 XX Claim 12; Page 10-17; 22pp; English.  
 PS  
 CC Transferrin binding protein 1 (Tbp1) (AAM04867) of Actinobacillus  
 CC pleuropneumoniae (AP) strain 1370, derived from strain Hpn-1  
 CC (ATCC 27088) of serotype 1, is probably a transmembrane protein  
 CC that serves as a channel for transport of iron across the outer  
 CC membrane. Its amino acid sequence was deduced from a gene  
 CC (AAT38071) obtd. from AP genomic DNA. Recombinant Tbp1, or its  
 CC antigenic fragments, can be produced in transformed host cells.  
 CC It is used to formulate vaccines against porcine pleuropneumonia,  
 CC to prepare antibodies (useful for serotherapy) and to prepare  
 CC diagnostic reagents.  
 CC  
 XX Sequence 931 AA;  
 SQ

Query Match 100.0%; Score 30; DB 17; Length 931;  
 Best Local Similarity 100.0%; Pred. No. 8.6e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLGGS 6  
 |||||  
 Db 151 gslggs 156

RESULT 13  
 AAM31346  
 ID AAM31346 standard; Protein; 1346 AA.  
 XX  
 AC AAM31346;  
 XX

DT 20-JUL-1998 (first entry)  
 XX  
 XX Rat tumour suppressor protein SSeCKS (active truncated form).  
 DE  
 XX  
 XX SSeCKS; tumour suppressor gene; rat; protein kinase C; mitosis;  
 KW cancer; malignancy; cell proliferation; Alzheimer's disease;  
 KW therapy.  
 KW  
 XX  
 OS Rattus sp.  
 XX  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 72  
 FT /note= "encoded by TCR"  
 FT Misc-difference 496  
 FT /note= "encoded by GTC"  
 FT Misc-difference 499  
 FT /note= "encoded by GTC"  
 FT Misc-difference 757  
 FT /note= "encoded by CCC"  
 FT Misc-difference 785  
 FT /note= "encoded by AGC"  
 FT Region 24..32  
 FT /note= "glycine-rich region"  
 FT Peptide 131..134  
 FT /note= "nuclear localisation signal"  
 FT Region 138..141  
 FT /note= "glycine-rich region"  
 FT Peptide 202..205  
 FT /note= "nuclear localisation signal"  
 FT Peptide 289..292  
 FT /note= "nuclear localisation signal"  
 FT Region 306..315  
 FT /note= "glycine-rich region"  
 FT

FT Peptide 368..371  
 FT /note= "nuclear localisation signal"  
 FT Region 1013..1035  
 FT /note= "zinc finger"  
 PN MO9740059-A1.  
 XX 30-OCT-1997.  
 XX 18-APR-1997: 97MO-US06830.  
 XX 18-JUN-1996: 96US-0665401.  
 PR 19-APR-1996: 96US-0635121.  
 XX  
 PA (GELM/) GELMAN I.  
 PA (JAKE/) JAKEN S.  
 PI Gelman I, Jaken S;  
 DR WPI: 1997-535770/49.  
 DR N-PSDB; AAV02301.  
 PT Tumour suppressor gene SSeCKS - used as a mitotic regulator, and  
 PT inhibitor of malignant phenotype  
 XX  
 PS Example 6: Fig 3A-G; 162pp; English.  
 XX  
 CC This polypeptide comprises an active truncated form of the novel  
 CC tumour suppressor protein SSeCKS that is a substrate of protein  
 CC kinase C and which acts as a negative regulatory of mitosis and as  
 CC an inhibitor of the transformed phenotype. Its amino acid sequence  
 CC was deduced from a partial cDNA clone (see AAV02301) obtained from  
 CC NIH3T3 cells. The full-length SSeCKS sequence (see AAW31347) is also  
 CC provided. Full-length SSeCKS nucleic acid (see AAV02301), as well as  
 CC homologous and hybridising nucleic acids are claimed, as are  
 CC isolated proteins encoded by such nucleic acids, vectors comprising  
 CC the nucleic acids, host cells, and methods of inhibiting the  
 CC expression of a transformed phenotype in a host cell by introducing  
 CC the nucleic acids. Introduction of a SSeCKS nucleic acid or gene  
 CC product into a host cell inhibits mitosis of the host cell,  
 CC allowing the treatment of diseases associated with disorders of  
 CC proliferation and/or with the expression of a malignant phenotype.  
 CC SSeCKS can also be used to treat or identify disorders of  
 CC cytoskeletal structure and cellular architecture (such as  
 CC Alzheimer's disease), and may be a marker for aberrancies in  
 CC fertility and/or nervous system development.  
 CC  
 SQ Sequence 1346 AA:  
 Query Match 100.0%; Score 30; DB 18; Length 1346;  
 Best Local Similarity 100.0%; Pident. No. 1.3e+03;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GSLGGS 6  
 Db 960 gslggs 965  
 RESULT 14  
 AAW31347  
 ID AAW31347 standard; Protein: 1596 AA.  
 AC AAW31347:  
 XX 20-JUL-1998 (first entry)  
 DE Rat tumour suppressor protein SSeCKS.  
 XX  
 KW SSeCKS; tumour suppressor gene; rat; protein kinase C; mitosis;  
 KM cancer; malignancy; cell proliferation; Alzheimer's disease;  
 XX therapy.

OS Rattus sp.  
 XX Key  
 FH Peptide  
 FT Location/Qualifiers  
 FT 1..10  
 FT /note= "myristylation site"  
 FT 410..418  
 FT /note= "glycine-rich region"  
 FT 517..520  
 FT /note= "nuclear localisation signal"  
 FT 524..527  
 FT /note= "glycine-rich region"  
 FT 677..680  
 FT /note= "nuclear localisation signal"  
 FT 757..760  
 FT /note= "nuclear localisation signal"  
 FT 1399..1421  
 FT /note= "nuclear localisation signal"  
 FT 279..307  
 FT /note= "zinc finger"  
 FT Modified-site  
 FT /note= "proposed protein kinase C phosphorylation  
 FT site"  
 FT 504..526  
 FT /note= "proposed protein kinase C phosphorylation  
 FT site"  
 FT Modified-site  
 FT /note= "proposed protein kinase C phosphorylation  
 FT site"  
 FT 592..614  
 FT /note= "proposed protein kinase C phosphorylation  
 FT site"  
 FT 741..766  
 FT /note= "proposed protein kinase C phosphorylation  
 FT site"  
 FT Modified-site  
 FT /note= "proposed protein kinase C phosphorylation  
 FT site"  
 FT 741..766  
 FT /note= "proposed protein kinase C phosphorylation  
 FT site"  
 PN MO9740059-A1.  
 XX 30-OCT-1997.  
 XX 18-APR-1997: 97MO-US06830.  
 XX 18-JUN-1996: 96US-0665401.  
 PR 19-APR-1996: 96US-0635121.  
 XX  
 PA (GELM/) GELMAN I.  
 PA (JAKE/) JAKEN S.  
 PI Gelman I, Jaken S;  
 DR WPI: 1997-535770/49.  
 DR N-PSDB; AAV02302.  
 PT Tumour suppressor gene SSeCKS - used as a mitotic regulator, and  
 PT inhibitor of malignant phenotype  
 XX  
 PS Claim 4: Fig 11A-L; 162pp; English.  
 XX  
 CC This polypeptide comprises the novel tumour suppressor protein  
 CC SSeCKS that is a substrate of protein kinase C and which acts as a  
 CC negative regulatory of mitosis and as an inhibitor of the  
 CC transformed phenotype. Its amino acid sequence was deduced from an  
 CC isolated nucleic acid molecule (see AAV02302). The SSeCKS nucleic  
 CC acid, as well as homologous and hybridising nucleic acids are  
 CC claimed, as are isolated proteins encoded by such nucleic acids, and  
 CC vectors comprising the nucleic acids, host cells, and methods of  
 CC inhibiting the expression of a transformed phenotype in a host cell  
 CC by introducing the nucleic acids. Introduction of a SSeCKS nucleic  
 CC acid or gene product into a host cell inhibits mitosis of the host  
 CC cell, allowing the treatment of diseases associated with disorders  
 CC of proliferation and/or with the expression of a malignant  
 CC phenotype. SSeCKS can also be used to treat or identify disorders  
 CC of cytoskeletal structure and cellular architecture (such as  
 CC Alzheimer's disease), and may be a marker for aberrancies in  
 CC fertility and/or nervous system development.  
 CC  
 SQ Sequence 1596 AA:



Query Match 100.0%; Score 30; DB 18; Length 1596;  
Best Local Similarity 100.0%; Pred. No. 1.5e+03;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLGGS 6  
|||||  
Db 1347 gslggs 1352

## RESULT 15

AAK57055  
ID AAK57055 standard; peptide; 21 AA.  
XX  
AC AAK57055;  
XX  
DT 15-MAR-1995 (first entry)  
XX  
DE HTLV peptide.  
XX  
KW HTLV1; HTLV2; HTLV-I; HTLV-II; diagnostic.  
XX  
OS HTLV.  
XX  
PN MO9418322-A.  
XX  
PD 18-AUG-1994.  
XX  
PE 02-FEB-1994; 94MO-US01170.  
PR 05-FEB-1993; 93US-0014153.  
XX  
PA (GENE-) GENELABS TECHNOLOGIES INC.  
XX (STRD ) UNIV LELAND STANFORD JUNIOR.  
PI Fong SKH, Goh C, Hadlock KG;  
XX  
DR WP1; 1994-279743/34.  
XX  
PT HTLV-I and -II peptide(s) and kits contg. them - used to diagnose  
XX infection and discriminate between HTLV-I and -II infection  
XX  
PS Disclosure; Page 74; 100pp; English.  
XX  
CC Novel HTLV1 and HTLV2 peptides are disclosed for use in diagnostic  
XX assays for detecting HTLV1 and HTLV2 infection in human sera. 2  
XX Peptides of the invention are given in AAK57055-56.  
XX  
SQ Sequence 21 AA;

Query Match 93.3%; Score 28; DB 15; Length 21;  
Best Local Similarity 83.3%; Pred. No. 39;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLGGS 6  
|||||  
Db 10 gsmggs 15

## RESULT 16

AAAG01209  
ID AAG01209 standard; Protein; 97 AA.  
XX  
AC AAG01209;  
XX  
DT 06-OCT-2000 (first entry)  
XX  
DE Human secreted protein, SEQ ID NO: 5290.  
XX  
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
XX gene therapy; chromosome mapping.  
OS Homo sapiens.

XX  
PN EP1033401-A2.  
XX  
PD 06-SEP-2000.  
XX

XX 21-FEB-2000; 2000EP-0200610.  
XX  
PR 26-FEB-1999; 99US-0122487.  
XX

XX (GEST ) GENSET.  
XX

XX Dumas Milne Edwards J, Duclert A, Giordano J;  
XX

XX WP1: 2000-500381/45.  
XX  
XX N-PSDB; AAC01215.  
XX

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX  
XX Claim 13; SEQ ID 5290; 71pp + CD-ROM; English.  
XX

XX The present sequence is a polypeptide encoded by one of a large number  
XX of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs  
XX were prepared from total human RNAs or polyA+ RNAs derived from 30  
XX different tissues. EST sequences usually correspond mainly to the 3'  
XX untranslated region (UTR) of the mRNA because they are often obtained  
XX from oligo-dT primed cDNA libraries. Such ESTs are not well suited for  
XX isolating cDNA sequences derived from the 5' ends of mRNAs and even in  
XX those cases where longer cDNA sequences have been obtained, the full 5'  
XX UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'  
XX ends and can therefore be used to obtain full length cDNAs and genomic  
XX DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and  
XX regulatory sequences and to design expression and secretion vectors.  
XX  
XX Sequence 97 AA;

Query Match 93.3%; Score 28; DB 21; Length 97;  
Best Local Similarity 83.3%; Pred. No. 1.9e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLGGS 6  
|||||  
Db 34 gslggs 39

## RESULT 17

AAAG74477  
ID AAG74477 standard; Protein; 119 AA.  
XX  
AC AAG74477;  
XX  
DT 03-SEP-2001 (first entry)  
XX  
DE Human colon cancer antigen protein SEQ ID NO:5241.  
XX  
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;  
XX colorectal carcinoma.  
XX  
OS Homo sapiens.  
XX  
XX WO200122920-A2.  
XX  
XX  
XX 05-APR-2001.  
XX  
XX 28-SEP-2000; 2000MO-US26524.  
XX  
XX 29-SEP-1999; 99US-0157137.  
XX  
XX 03-NOV-1999; 99US-0163280.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.

XX PI Ruben SM, Barash SC, Birse CE, Rosen CA;  
XX DR WPI: 2001-235357/24.  
XX DR N-PSDB; AAH33908.  
XX PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
PT useful for preventing, diagnosing and/or treating colorectal cancers -  
XX PS  
XX PS Claim 11: Page 6914-6915; 9803pp; English.  
XX CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon  
CC cancer-associated nucleic acid molecules (N) and proteins (P), where  
CC the proteins are collectively known as colon cancer antigens. The colon  
CC cancer antigens have cytostatic activity and can be used in gene  
CC therapy and vaccine production. N and P may be used in the prevention,  
CC diagnosis and treatment of diseases associated with inappropriate P  
CC expression. For example, N and P may be used to treat disorders  
CC associated with decreased expression by rectifying mutations or deletions  
CC in a patient's genome that affect the activity of P by expressing  
CC inactive proteins or to supplement the patient's own production of P.  
CC Additionally, N may be used to produce the colon cancer-associated Ps,  
CC by inserting the nucleic acids into a host cell and culturing the cell  
CC to express the proteins. N and P can be used in the prevention, diagnosis  
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
CC and AAB77789 represent sequences used in the exemplification of the  
CC present invention.  
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
CC missing at time of publication, meaning no sequences are present for  
CC SEQ ID NO:1027 to 1052, 7921 and 7922.  
SQ Sequence 119 AA;  
  
Query Match 93.3%; Score 28; DB 22; Length 119;  
Best Local Similarity 83.3%; Pred. NO. 2.4e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 GSLGGS 6  
||:||||  
Db 93 gslggs 98  
  
RESULT 18  
AAW57774  
ID AAW5774 standard; Protein: 633 AA.  
XX  
XX AAW5774;  
XX  
XX 02-DEC-1998 (first entry)  
XX  
XX Amino acid sequence of lepidopteran-active HD525 toxin.  
XX  
XX HD525 toxin; PCR: primer; amplification; Bacillus thuringiensis; probe;  
KM Lepidoptera; pest; pesticide; Ostrinia nubilalis; Heliothis virescens;  
KM Helicoverpa zea; hybridisation.  
XX  
XX Bacillus thuringiensis.  
XX  
XX WO9840490-A1.  
XX  
XX 17-SEP-1998.  
XX  
XX 13-MAR-1998; 98WO-0505081.  
XX  
XX 13-MAR-1997; 97US-0040512.  
XX  
XX (MYCO ) MYCOGEN CORP.  
XX  
XX Muller-Cohn J, Narva KE, Schnepf HE;  
XX WPI: 1998-506734/43.  
XX DR N-PSDB; AAV52611.

XX  
PT New Insecticidal Bacillus thuringiensis toxins - useful for  
PT controlling lepidopteran pests, especially Ostrinia nubilalis,  
PT Heliothis virescens and Helicoverpa zea  
XX  
XX PS  
XX PS Claim 14: Pages 32-34; 50pp; English.  
XX  
XX This is the amino acid sequence of a novel Bacillus thuringiensis toxin  
CC used in the method of the invention, to control lepidopteran pests.  
CC The new toxins are useful as pesticides, especially for the control of  
CC Ostrinia nubilalis, Heliothis virescens, and Helicoverpa zea. The  
CC polynucleotide coding sequences are useful for recombinant expression  
CC of the toxins and the primers, together with probes derived from the  
CC new sequences, are useful for the identification and characterisation  
CC of novel genes that encode pesticidal toxins.  
SQ Sequence 633 AA;  
  
Query Match 93.3%; Score 28; DB 19; Length 633;  
Best Local Similarity 83.3%; Pred. NO. 1.4e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 GSLGGS 6  
||:||||  
Db 349 gslggs 354  
  
RESULT 19  
AAW57775  
ID AAW5775 standard; Protein: 633 AA.  
XX  
XX AAW5775;  
XX  
XX 02-DEC-1998 (first entry)  
XX  
XX Amino acid sequence of lepidopteran-active 8612 toxin.  
XX  
XX 8612 toxin; PCR: primer; amplification; Bacillus thuringiensis; probe;  
KM Lepidoptera; pest; pesticide; Ostrinia nubilalis; Heliothis virescens;  
KM Helicoverpa zea; hybridisation.  
XX  
XX Bacillus thuringiensis.  
XX  
XX WO9840490-A1.  
XX  
XX 17-SEP-1998.  
XX  
XX 13-MAR-1998; 98WO-0505081.  
XX  
XX 13-MAR-1997; 97US-0040512.  
XX  
XX (MYCO ) MYCOGEN CORP.  
XX  
XX Muller-Cohn J, Narva KE, Schnepf HE;  
XX WPI: 1998-506734/43.  
XX DR N-PSDB; AAV52612.  
XX  
XX New insecticidal Bacillus thuringiensis toxins - useful for  
PT controlling lepidopteran pests, especially Ostrinia nubilalis,  
PT Heliothis virescens and Helicoverpa zea  
XX  
XX PS  
XX PS Claim 5: Pages 36-38; 50pp; English.  
XX  
XX This is the amino acid sequence of a novel Bacillus thuringiensis toxin  
CC used in the method of the invention, to control lepidopteran pests.  
CC The new toxins are useful as pesticides, especially for the control of  
CC Ostrinia nubilalis, Heliothis virescens, and Helicoverpa zea. The  
CC polynucleotide coding sequences are useful for recombinant expression  
CC of the toxins and the primers, together with probes derived from the  
CC new sequences, are useful for the identification and characterisation  
CC of novel genes that encode pesticidal toxins.

XX Sequence 633 AA;  
SQ  
Query Match 93.3%; Score 28; DB 19; Length 633;  
Best Local Similarity 83.3%; Pred. No. 1.4e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GSIGS 6  
||:||||  
Db 349 gsiggs 354  
RESULT 20  
AAG50497  
ID AAG50497 standard; Protein; 1892 AA.  
XX AAG50497;  
AC  
XX AAG50497;  
DT 18-OCT-2000 (first entry)  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 64001.  
XX  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
XX Arabidopsis thaliana.  
OS  
PN EP1033405-A2.  
PD 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
PF  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0122180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134376.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145088.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145291.  
PR 27-JUL-1999; 99US-0145919.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.

PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157717.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158222.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159295.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159754.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
Query Match 93.3%; Score 28; DB 21; Length 1892;  
Best Local Similarity 83.3%; Pred. No. 4.4e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GSUGS 6  
11:111  
Db 650 gs19gs 655  
RESULT 21  
AAG50496  
ID AAG50496 standard. Protein: 1940 AA.  
XX  
XX AAG50496;  
XX  
XX 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 64000.  
XX  
XX protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
XX EP1033405-A2.  
XX  
XX 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
XX 05-MAR-1999; 99US-0123180.  
XX 09-MAR-1999; 99US-0123548.  
XX 23-MAR-1999; 99US-0125788.  
XX 25-MAR-1999; 99US-0126264.  
XX 29-MAR-1999; 99US-0126785.  
XX 01-APR-1999; 99US-0127462.  
XX 06-APR-1999; 99US-0128234.  
XX 08-APR-1999; 99US-0128714.  
XX 16-APR-1999; 99US-0129845.  
XX 19-APR-1999; 99US-0130077.  
XX 21-APR-1999; 99US-0130449.  
XX 23-APR-1999; 99US-0130510.  
XX 23-APR-1999; 99US-0130691.  
XX 28-APR-1999; 99US-0131449.  
XX 30-APR-1999; 99US-0132048.  
XX 30-APR-1999; 99US-0132407.  
XX 04-MAY-1999; 99US-0132484.  
XX 05-MAY-1999; 99US-0132485.  
XX 06-MAY-1999; 99US-0132486.  
XX 06-MAY-1999; 99US-0132487.  
XX 07-MAY-1999; 99US-0132863.  
XX 11-MAY-1999; 99US-0134256.  
XX 14-MAY-1999; 99US-0134219.  
XX 14-MAY-1999; 99US-0134221.  
XX 14-MAY-1999; 99US-0134370.  
XX 18-MAY-1999; 99US-0134378.  
XX 19-MAY-1999; 99US-0134941.  
XX 20-MAY-1999; 99US-0135124.  
XX 21-MAY-1999; 99US-0135353.  
XX 21-MAY-1999; 99US-0135629.  
XX 25-MAY-1999; 99US-0136021.  
XX 25-MAY-1999; 99US-0136782.  
XX 27-MAY-1999; 99US-0136782.  
XX 28-MAY-1999; 99US-0137222.  
XX 03-JUN-1999; 99US-0137528.  
XX 04-JUN-1999; 99US-0137502.  
XX 07-JUN-1999; 99US-0137724.  
XX 08-JUN-1999; 99US-0138094.  
XX 10-JUN-1999; 99US-0138540.  
XX 10-JUN-1999; 99US-0138847.  
XX 14-JUN-1999; 99US-0139119.  
XX 16-JUN-1999; 99US-0139452.  
XX 16-JUN-1999; 99US-0139453.

PR	17-JUN-1999;	99US-0139492.	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139454.	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139455.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139456.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139457.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139460.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139462.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139463.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139750.	PR	27-AUG-1999;	99US-0151065.
PR	21-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151066.
PR	22-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151080.
PR	22-JUN-1999;	99US-0139899.	PR	30-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140353.	PR	31-AUG-1999;	99US-0151438.
PR	23-JUN-1999;	99US-0140354.	PR	01-SEP-1999;	99US-0151930.
PR	24-JUN-1999;	99US-0140695.	PR	07-SEP-1999;	99US-0152363.
PR	28-JUN-1999;	99US-0140823.	PR	10-SEP-1999;	99US-0153070.
PR	29-JUN-1999;	99US-0140991.	PR	13-SEP-1999;	99US-0153758.
PR	30-JUN-1999;	99US-0141287.	PR	15-SEP-1999;	99US-0154018.
PR	01-JUL-1999;	99US-0141842.	PR	16-SEP-1999;	99US-0154033.
PR	01-JUL-1999;	99US-0142153.	PR	20-SEP-1999;	99US-0154779.
PR	02-JUL-1999;	99US-0142055.	PR	22-SEP-1999;	99US-0155139.
PR	06-JUL-1999;	99US-0142390.	PR	23-SEP-1999;	99US-0155486.
PR	08-JUL-1999;	99US-0142803.	PR	24-SEP-1999;	99US-0155659.
PR	09-JUL-1999;	99US-0142920.	PR	28-SEP-1999;	99US-0156458.
PR	12-JUL-1999;	99US-0142977.	PR	29-SEP-1999;	99US-0156596.
PR	13-JUL-1999;	99US-0143542.	PR	04-OCT-1999;	99US-0157117.
PR	14-JUL-1999;	99US-0143624.	PR	05-OCT-1999;	99US-0157153.
PR	15-JUL-1999;	99US-0144005.	PR	06-OCT-1999;	99US-0157865.
PR	16-JUL-1999;	99US-0144085.	PR	07-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144086.	PR	08-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144332.	PR	12-OCT-1999;	99US-0158363.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144333.	PR	14-OCT-1999;	99US-0159330.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159330.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159637.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159638.
PR	21-JUL-1999;	99US-0144884.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160776.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160815.
PR	22-JUL-1999;	99US-0145192.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0160981.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160989.
PR	23-JUL-1999;	99US-0145224.	PR	25-OCT-1999;	99US-0161404.
PR	26-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161406.
PR	27-JUL-1999;	99US-0145918.	PR	26-OCT-1999;	99US-0161359.
PR	27-JUL-1999;	99US-0145919.	PR	26-OCT-1999;	99US-0161360.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146386.	PR	28-OCT-1999;	99US-0161920.
PR	02-AUG-1999;	99US-0146388.	PR	28-OCT-1999;	99US-0161923.
PR	03-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161993.
PR	04-AUG-1999;	99US-0147204.	PR	29-OCT-1999;	99US-0162142.
PR	04-AUG-1999;	99US-0147302.			
PR	05-AUG-1999;	99US-0147192.			
PR	05-AUG-1999;	99US-0147260.			
PR	06-AUG-1999;	99US-0147303.			
PR	06-AUG-1999;	99US-0147416.			
PR	09-AUG-1999;	99US-0147493.			
PR	09-AUG-1999;	99US-0147935.			
PR	10-AUG-1999;	99US-0148177.			
PR	11-AUG-1999;	99US-0148319.			
PR	12-AUG-1999;	99US-0148341.			
PR	13-AUG-1999;	99US-0148565.			

Query Match

Best Local Similarity

Matches

5; Conservative

93.3%;

83.3%;

1;

Score 28; DB 21; Length 1940;

Pred. No. 4.5e+03;

Mismatches 0;

Indels 0;

Gaps 0;

QY

1

GSIGGS

6

11:111

703

Db

698

gsigs

703

RESULT

22

AG50495  
ID AG50495 standard; Protein: 1944 AA.  
XX  
AC AG50495;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 63999.  
XX  
KW Protein identification: signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126284.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 15-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134421.  
PR 18-MAY-1999; 99US-0134370.  
PR 19-MAY-1999; 99US-0134768.  
PR 20-MAY-1999; 99US-0134941.  
PR 21-MAY-1999; 99US-0135124.  
PR 24-MAY-1999; 99US-0135353.  
PR 25-MAY-1999; 99US-0135629.  
PR 27-MAY-1999; 99US-0136021.  
PR 28-MAY-1999; 99US-0136392.  
PR 01-JUN-1999; 99US-0136782.  
PR 03-JUN-1999; 99US-0137222.  
PR 04-JUN-1999; 99US-0137528.  
PR 07-JUN-1999; 99US-0137502.  
PR 08-JUN-1999; 99US-0137724.  
PR 10-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 14-JUN-1999; 99US-0138847.  
PR 16-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 24-JUN-1999; 99US-0140354.  
PR 28-JUN-1999; 99US-0140685.  
PR 28-JUN-1999; 99US-0140823.  
PR 30-JUN-1999; 99US-0140991.  
PR 01-JUL-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145513.  
PR 27-JUL-1999; 99US-0145518.  
PR 27-JUL-1999; 99US-0145519.  
PR 28-JUL-1999; 99US-0145591.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.

PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0158293.  
PR 13-OCT-1999; 99US-0158294.  
PR 13-OCT-1999; 99US-0158295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 93.3%; Score 28; DB 21; Length 1944;  
Best Local Similarity 83.3%; Pred. No. 4.5e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSLGGS 6  
||:||||  
DB 702 gslggs 707

RESULT 23  
AAV04755  
ID AAV04755 standard; Protein; 67 AA.  
XX  
AC AAV04755;  
XX  
DT 06-JUL-1999 (first entry)  
XX

DE Mycobacterium species protein sequence 244.  
XX  
XX Secreted protein; Mycobacterium; primer; PCR; amplification; probe;  
KW hybridisation; detection; vaccine; immunisation; infection.  
XX  
OS Mycobacterium sp.  
XX  
XX WO9909186-A2.  
XX  
PD 25-FEB-1999.  
XX  
PF 14-AUG-1998; 98WO-FR01813.  
XX  
PR 11-SEP-1997; 97FR-0011325.  
PR 14-AUG-1997; 97FR-0010404.  
XX  
PA (INSP ) INST PASTEUR.  
XX  
XX Gicquel B, Lim EM, Pellicle V, Portnoi D, Goguet de la Salmoniere Y;  
PI Guigueno A;  
PI  
XX  
DR WPI: 1999-181045/15.  
DR N-PSDB: AAX34007.  
XX  
XX  
XX Mycobacterial DNA vectors containing reporter constructs - for  
PT identifying coding or promoter sequences involved in  
PT infection-associated protein expression  
XX  
XX Claim 32; Fig 2; 309pp; French.  
XX  
XX Sequences AAV04742-Y05000 and AAV07201-Y07204 represent secreted  
CC proteins from various Mycobacterium species microorganisms. The  
CC encoding nucleotide sequences can be used as primers and probes for  
CC methods for detecting and identifying mycobacteria, especially belonging  
CC to the M. tuberculosis complex. The encoded proteins can be used in  
CC vaccines for immunisation against a bacterial or viral infection.  
XX  
SQ Sequence 67 AA;

Query Match 90.0%; Score 27; DB 20; Length 67;  
Best Local Similarity 83.3%; Pred. No. 2e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSLGGS 6  
||:||||  
DB 59 gsvggs 64

RESULT 24  
AAG27140  
ID AAG27140 standard; Protein; 91 AA.  
XX  
AC AAG27140;  
XX

DT 17-OCT-2000 (first entry)  
XX

DE Zea mays protein fragment SEQ ID NO: 31861.  
XX

KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence; corn.  
XX

OS Zea mays subsp. mays.  
XX

PN EP1033405-A2.  
XX

PD 06-SEP-2000.  
XX

PF 25-FEB-2000; 2000EP-0301439.  
XX

PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135134.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137232.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137529.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138004.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139730.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140635.  
PR 26-JUN-1999; 99US-0140683.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143562.  
PR 14-JUL-1999; 99US-0143624.  
  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144382.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148665.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149375.  
PR 17-AUG-1999; 99US-0149178.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155559.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.



PR 06-OCT-1999; 9905-0157865.  
PR 07-OCT-1999; 9905-0158029.  
PR 08-OCT-1999; 9905-0158232.  
PR 12-OCT-1999; 9905-0158369.  
PR 13-OCT-1999; 9905-0159293.  
PR 13-OCT-1999; 9905-0159294.  
PR 13-OCT-1999; 9905-0159295.  
PR 14-OCT-1999; 9905-0159329.  
PR 14-OCT-1999; 9905-0159330.  
PR 14-OCT-1999; 9905-0159331.  
PR 14-OCT-1999; 9905-0159637.  
PR 18-OCT-1999; 9905-0159638.  
PR 21-OCT-1999; 9905-0159584.  
PR 21-OCT-1999; 9905-0160741.  
PR 21-OCT-1999; 9905-0160767.  
PR 21-OCT-1999; 9905-0160768.  
PR 21-OCT-1999; 9905-0160770.  
PR 21-OCT-1999; 9905-0160814.  
PR 21-OCT-1999; 9905-0160815.  
PR 22-OCT-1999; 9905-0160980.  
PR 22-OCT-1999; 9905-0160981.  
PR 22-OCT-1999; 9905-0160989.  
PR 25-OCT-1999; 9905-0161404.  
PR 25-OCT-1999; 9905-0161405.  
PR 25-OCT-1999; 9905-0161406.  
PR 26-OCT-1999; 9905-0161359.  
PR 26-OCT-1999; 9905-0161360.  
PR 26-OCT-1999; 9905-0161361.  
PR 28-OCT-1999; 9905-0161920.  
PR 28-OCT-1999; 9905-0161992.  
PR 28-OCT-1999; 9905-0161993.  
PR 29-OCT-1999; 9905-0162142.

Query Match 90.0%; Score 27; DB 21; Length 91;  
Best Local Similarity 83.3%; Pred. No. 2.8e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSIGGS 6  
Db 1 gslgga 6

RESULT 25  
ID AAB45378 standard; Protein; 121 AA.  
AC AAB45378;  
XX  
DT 14-FEB-2001 (first entry)

DE Human secreted protein sequence encoded by gene 38 SEQ ID NO:130.  
XX  
KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;  
KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;  
KW cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;  
KW fungicide; ophthalmological; vulnerary; gene therapy; autoimmune disease;  
KW hyperproliferative disorder; cardiovascular disorder; angiogenesis;  
KW cerebrovascular disorder; nervous system disorder; infection; skin aging;  
KW ocular disorder; wound healing; food additive; preservative.

OS Homo sapiens.  
XX  
XX WO200061628-A1.  
XX  
PD 19-OCT-2000.  
XX  
PF 06-APR-2000; 2000MO-US09070.  
XX  
PR 09-APR-1999; 9905-0128695.  
XX  
PR 14-JAN-2000; 200005-0176052.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX

PI Rosen CA, Ruben SM, Komatsoulis G;  
XX  
DR WPI, 2000-619228/59.  
XX  
PT New nucleic acid molecules encoding 49 human secreted proteins for  
PT diagnosing, preventing, treating or ameliorating medical conditions and  
PT used as food additives or preservatives -  
XX

PS Disclosure; Page 449; 454pp; English.

CC The polynucleotide sequences given in AAC81086 to AAC81134 encode the  
CC human secreted proteins given in AAB45308 to AAB45356. AAB45357 to  
CC AAB45384 represent human secreted polypeptide sequences and proteins  
CC homologous to them, which are given in the exemplification of the present  
CC invention. Human secreted proteins have activities based on the tissues  
CC and cells the genes are expressed in. Examples of activities include:  
CC antirheumatic; immunosuppressive; antirheumatic; antiproliferative;  
CC cytostatic; cardiant; vasotropic; cerebroprotective; neurotropic;  
CC neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
CC and vulnerary. The polynucleotides and polypeptides can be used to  
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used  
CC in diagnosing a pathological condition or susceptibility to a  
CC pathological condition. Disorders which are diagnosed or treated include  
CC autoimmune diseases, hyperproliferative disorders, cardiovascular  
CC disorders, cerebrovascular disorders, angioneuromas, nervous system  
CC disorders. The polypeptides can also be used to aid wound healing and  
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to  
CC maintain organs before transplantation, for supporting cell culture of  
CC primary tissues, to regenerate tissues and in chemotaxis. The  
CC polypeptides can also be used as a food additive or preservative to  
CC increase or decrease storage capabilities, fat content, lipid, protein,  
CC carbohydrate, vitamins, minerals, cofactors and other nutritional  
CC components. AAC81077 to AAC81085 and AAB45307 represent sequences used in  
CC the exemplification of the present invention.  
XX

Sequence 121 AA:

Query Match 90.0%; Score 27; DB 21; Length 121;  
Best Local Similarity 83.3%; Pred. No. 3.8e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSIGGS 6  
Db 36 gslgga 41

RESULT 26  
ID AAY59979 standard; Protein; 126 AA.  
XX  
AC AAY59979;

DT 31-JAN-2000 (first entry)  
XX  
DE Human endometrium tumour EST encoded protein 39.

XX  
KW Endometrium; human; tumour; cancer; anticancer; cytostatic; EST;  
KW treatment; uterine; gene therapy; expressed sequence tag.

OS Homo sapiens.

XX  
XX DE19817948-A1.  
XX  
PD 21-OCT-1999.  
XX  
PF 17-APR-1998; 98DE-1017948.  
XX  
PR 17-APR-1998; 98DE-1017948.  
XX  
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.  
XX

XX Rosenchal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;  
 XX WPI: 1999-591957/51.  
 DR N-PSDB; AA241993.  
 XX  
 PT New nucleic acid sequences expressed in uterine cancer tissues, and  
 PT derived polypeptides, for treatment of uterine and endometrial cancer  
 PT and identification of therapeutic agents -  
 XX  
 PS Claim 23; Page 290; 444pp; German.  
 XX  
 CC This invention describes novel human nucleic acid (cDNA) sequences (A),  
 CC that are highly expressed in uterine tumour tissue and which have  
 CC anticancer and cytostatic activity. (A) are used (i) for recombinant  
 CC expression of polypeptides (B) and (ii) to isolate complete genes. (B)  
 CC are used (i) to identify agents suitable for treatment of uterine or  
 CC endometrial cancer; (ii) directly for treating these forms of cancer  
 CC (including expression from gene therapy vectors) and (iii) for  
 CC generation of specific antibodies. (A) are identified by assembling ESTs  
 CC (expressed sequence tags) from a particular tissue type before comparison  
 CC of expression patterns. This allows a significantly longer fragment of  
 CC the gene to be revealed, so should reduce the number of failures  
 CC associated with the fact that ESTs from different libraries may represent  
 CC different parts of the same unknown gene, distorting the estimated  
 CC frequency of occurrence in a particular tissue. AA59941-Y60328 represent  
 CC protein fragments encoded by the human endometrium tumour cDNA library  
 CC derived EST fragments represented in AA241981-242121.  
 CC  
 SO Sequence 126 AA;  
 XX

Query Match 90.0%; Score 27; DB 20; Length 126;  
 Best Local Similarity 83.3%; Pred. No. 4e+02;  
 Matches 5; Conservative 1; Mismatches 0; Gaps 0;  
 OY 1 GSLGGS 6  
 II:III  
 DB 52 gtlggs 57

RESULT 27  
 AAB95004  
 ID AAB95004 standard; Protein: 126 AA.  
 XX  
 AC AAB95004;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human protein sequence SEQ ID NO:16665.  
 XX  
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-0116126.  
 XX  
 PR 29-JUL-1999; 99JP-0248036.  
 XX  
 PR 27-AUG-1999; 99JP-0300253.  
 XX  
 PR 11-JAN-2000; 2000JP-0118776.  
 XX  
 PR 02-MAY-2000; 2000JP-0183767.  
 XX  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 DR WPI: 2001-318749/34.

XX  
 PT primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 XX  
 PS Claim 8; SEQ ID 16665; 2537pp + CD ROM; English.  
 XX  
 CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 CC  
 SO Sequence 126 AA;  
 XX

Query Match 90.0%; Score 27; DB 22; Length 126;  
 Best Local Similarity 83.3%; Pred. No. 4e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 GSLGGS 6  
 II:III  
 DB 101 gsvggs 106

RESULT 28  
 AAW53009  
 ID AAW53009 standard; Protein: 163 AA.  
 XX  
 AC AAW53009;  
 XX  
 DT 03-AUG-1998 (first entry)  
 XX  
 DE Mus musculus I-mfc protein.  
 XX  
 KW I-mfc; inhibitor of MYO D family; treatment; diagnosis; myogenesis;  
 KW defects; abnormal development; disease; cleidocranial dysplasia;  
 KW CDD; rhabdomyosarcoma; muscle tissue cancer.  
 XX  
 OS Mus musculus.  
 XX  
 PN W09808860-A1.  
 XX  
 PD 05-MAR-1998.  
 XX  
 PF 21-AUG-1997; 97WO-US14780.  
 XX  
 PR 27-AUG-1996; 96US-0704931.  
 XX  
 PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.  
 XX  
 PI Chen CMA, Groulaine M, Kraut N, Weintraub H;  
 XX  
 DR WPI: 1998-179377/16.  
 DR N-PSDB; AAV21284.

XX Inhibitor of Myod family proteins - useful for, e.g., treatment and  
PT diagnosis of defects in myogenesis responsible for abnormal  
PT development  
XX  
PS Disclosure: Pages 79-80; 92pp; English.  
XX  
XX The sequence is that of murine I-mfc (inhibitor of MyoD family)  
CC protein. Probes from the gene sequence can be used for determining the  
CC presence of an I-mfc protein or analogue, or for detecting I-mfc  
CC agonist activity in a test substance. The sequence can be used  
CC to treat, model and diagnose defects in myogenesis responsible  
CC for abnormal development and disease conditions in mammals,  
CC including humans. These include the human disease cleidocranial  
CC dysplasia (CDD), rhabdomyosarcomas and other cancers affecting  
CC muscle tissue in mammals.  
CC  
XX  
SQ Sequence 163 AA:  
  
Query Match 90.0%; Score 27; DB 19; Length 163;  
Best Local Similarity 83.3%; Pred. No. 5.2e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 GSIGGS 6  
1:||||  
Db 120 galsgs 125  
  
RESULT 29  
AAV35432  
ID AAV35432 standard; Protein; 173 AA.  
XX  
AC AAV35432;  
XX  
XX 13-SEP-1999 (first entry)  
XX  
DE Chlamydia pneumoniae transmembrane protein sequence.  
XX  
XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
KW vaccine; neutralising epitope.  
XX  
OS Chlamydia pneumoniae.  
XX  
PN WO927105-A2.  
XX  
PD 03-JUN-1999.  
XX  
PF 20-NOV-1998; 98MO-IB01890.  
XX  
PR 04-NOV-1998; 98US-0107078.  
PR 21-NOV-1997; 97FR-0014673.  
XX  
PA (GEST ) GENSET.  
PI Griffiths R;  
DR WPI; 1999-357842/30.  
XX  
PT Genome sequence of Chlamydia pneumoniae  
XX  
PS Page 1213; Disclosure: 1912pp; English.  
XX  
XX AAV34584-Y35879 represent the proteins encoded by all the open reading  
CC frames in the complete genome (see AAV31990) of Chlamydia pneumoniae.  
CC C. pneumoniae causes respiratory disease such as pneumonia and  
CC bronchitis and is thought to be a contributing factor in heart  
CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema  
CC nodosum or pharyngitis. The polypeptides encoded by the open reading  
CC frames of the C. pneumoniae genome (see AAV34584-Y35879) can be used in  
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae  
CC nucleotide sequences can also be used as immunogenic compositions,

CC especially where the vector directs the expression of a neutralising  
CC epitope of C. pneumoniae.  
XX  
XX  
SQ Sequence 173 AA:  
  
Query Match 90.0%; Score 27; DB 20; Length 173;  
Best Local Similarity 83.3%; Pred. No. 5.5e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 GSIGGS 6  
1:||||  
Db 46 gtlggs 51  
  
RESULT 30  
AAG24867  
ID AAG24867 standard; Protein; 168 AA.  
XX  
AC AAG24867;  
XX  
XX 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 28703.  
XX  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
XX Arabidopsis thaliana.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 19-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 17-JUN-1999; 99US-0139453.  
 PR 18-JUN-1999; 99US-0139492.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139817.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 23-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140695.  
 PR 28-JUN-1999; 99US-0140823.  
 PR 29-JUN-1999; 99US-0140991.  
 PR 30-JUN-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141842.  
 PR 01-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0142055.  
 PR 06-JUL-1999; 99US-0142390.  
 PR 08-JUL-1999; 99US-0142803.  
 PR 09-JUL-1999; 99US-0142920.  
 PR 12-JUL-1999; 99US-0142920.  
 PR 13-JUL-1999; 99US-0143542.  
 PR 14-JUL-1999; 99US-0143624.  
 PR 15-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 16-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 19-JUL-1999; 99US-0144335.  
 PR 20-JUL-1999; 99US-0144352.  
 PR 20-JUL-1999; 99US-0144632.  
 PR 20-JUL-1999; 99US-0144884.  
 PR 21-JUL-1999; 99US-0144814.  
 PR 21-JUL-1999; 99US-0145086.  
 PR 21-JUL-1999; 99US-0145088.  
 PR 22-JUL-1999; 99US-0145085.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 22-JUL-1999; 99US-0145129.  
 PR 23-JUL-1999; 99US-0145145.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 23-JUL-1999; 99US-0145224.  
 PR 26-JUL-1999; 99US-0145276.  
 PR 27-JUL-1999; 99US-0145913.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 28-JUL-1999; 99US-0145951.  
 PR 02-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 02-AUG-1999; 99US-0146389.  
 PR 03-AUG-1999; 99US-0147038.  
 PR 04-AUG-1999; 99US-0147204.  
 PR 04-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147192.

PR 05-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147935.  
 PR 10-AUG-1999; 99US-0148171.  
 PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151085.  
 PR 27-AUG-1999; 99US-0151086.  
 PR 27-AUG-1999; 99US-0151086.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155486.  
 PR 28-SEP-1999; 99US-0155659.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157117.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 26-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 28-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match

90.0%; Score 27; DB 21; Length 188;

Best Local Similarity 83.3%; Pred. No. 6e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSLGGS 6  
|:||||  
Db 72 galsgs 77

RESULT 31  
AAB39461

ID AAB39461 standard; Protein; 190 AA.

XX AAB39461;

DT 02-FEB-2001 (first entry)

DE Human secreted protein sequence encoded by gene 18 SEQ ID NO:120.

XX  
XX Human; secreted protein; diagnosis; cytostatic; immunosuppressive;  
KW neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic;  
KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; antidiabetic;  
KW antidiabetic; antifungal; antiparasitic; cardiant; gene therapy;  
KW food additive; preservative; chromosome identification; cancer;  
KW female reproductive system disorder; immune disorder; wound healing;  
KW cardiovascular disorder; neurological disease; infectious disease;  
KW infection.

XX Homo sapiens.

PN MO200058340-A2.

PD 05-OCT-2000.

PF 23-MAR-2000; 2000MO-US07724.

PR 26-MAR-1999; 99US-0126510.

PR 07-JAN-2000; 2000US-0174850.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM, Komatsoulis G;

XX MPI; 2000-594638/56.

PT Fifty nucleic acid molecules encoding human secreted proteins, useful  
PT in the prevention, treatment and diagnosis of cancer, immune disorders,  
PT cardiovascular disorders and neurological diseases -

PS Disclosure; Page 381; 391pp; English.

CC The polynucleotide sequences given in AAC74337 to AAC74386 encode the  
CC human secreted proteins given in AAB394402 to AAB39451. AAB39452 to  
CC AAB39484 represent human secreted polypeptide sequences and proteins  
CC homologous to them, which are given in the exemplification of the present  
CC invention. Human secreted proteins have activities based on the tissues  
CC and cells the genes are expressed in. Example of activities include:  
CC cytostatic; immunosuppressive; neurotropic; neuroprotective; antiviral;  
CC antiallergic; hepatotropic; antidiabetic; antidiabetic; antidiabetic;  
CC vulnary; anticonvulsant; antibacterial; antidiabetic; antidiabetic;  
CC cardiant. The polynucleotides and polypeptides are useful for preventing,  
CC treating or ameliorating a medical condition in e.g. humans, mice,  
CC rabbits, goats, horses, cats, dogs, chickens or sheep. The polypeptides  
CC can also be used as a food additive or preservative to increase or  
CC decrease storage capabilities. The polynucleotide are useful for  
CC chromosome identification. They are also useful as probes for diagnosing  
CC a disorder related to the female reproductive system, particularly breast  
CC and/or ovarian cancer. They are also useful in the gene therapy of breast  
CC antibodies, agonists and antagonists are useful in the diagnosis,  
CC treatment and prevention of: (a) cancer; (b) immune disorders; (c)  
CC cardiovascular disorders; (d) wound healing; (e) neurological diseases;  
CC and (f) infectious diseases such as viral, bacterial, fungal and  
CC parasitic infections. AAC74328 to AAC74336 and AAB39401 represent

CC sequences used in the exemplification of the present invention.

XX SQ Sequence 190 AA;

Query Match 90.0%; Score 27; DB 21; Length 190;

Best Local Similarity 83.3%; Pred. No. 6.1e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSLGGS 6  
|:||||  
Db 95 galsga 100

RESULT 32  
AAG14530

ID AAG14530 standard; Protein; 192 AA.

XX AAG14530;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 14426.

XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0128845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132485.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139753.  
PR 18-JUN-1999; 99US-0139817.  
PR 21-JUN-1999; 99US-0139899.  
PR 22-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.

PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149920.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150366.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154179.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 18-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 23-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match  
Best Local Similarity

90.0%;  
83.3%;

Score 27; DB 21; Length 192;  
Pred. No. 6.2e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GSIGS 6  
Db 74 galsgs 79

## RESULT 33

AAW81735  
ID AAW81735 standard; Protein: 196 AA.

XX AAW81735;

DT 27-JAN-1999 (first entry)

DE M. tuberculosis immunogenic polypeptide XP27 protein #2.

KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;

KV vaccine; pharmaceutical; infection; diagnosis.

XX Mycobacterium tuberculosis.

XX WO9816646-A2.

PD 23-APR-1998.

PF 07-OCT-1997; 97WO-US18293.

XX 13-MAR-1997; 97US-0818112.

PR 11-OCT-1996; 96US-0730510.

PA (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;

PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

DR N-PSDB; AAV64530.

XX WPI; 1998-261042/23.

XX N-PSDB; AAV64530.

PS Example 3c; Page 182; 230pp; English.

CC This sequence represents an immunogenic portion of a soluble  
CC Mycobacterium tuberculosis (MT) antigen which can be used in a method  
CC for inducing protective immunity against tuberculosis (TB). This sequence  
CC can be formulated into vaccines and/or pharmaceutical compositions for  
CC immunising against M. tuberculosis infection or may be used for the  
CC diagnosis of tuberculosis.

SQ Sequence 196 AA;

Query Match 90.0%; Score 27; DB 19; Length 196;

Best Local Similarity 83.3%; Pred. No. 6.3e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GSIGS 6  
Db 190 gsvgs 195

## RESULT 34

AAW64370  
ID AAW64370 standard; Protein: 196 AA.

XX AAW64370;

DT 09-NOV-1998 (first entry)

DE Mycobacterium tuberculosis antigen XP36.

XX Tuberculosis; infection; diagnosis; antigen; XP36.

OS Mycobacterium tuberculosis strain Erdman.

XX WO9816645-A2.

XX 23-APR-1998.

PF 07-OCT-1997; 97WO-US18214.

XX 13-MAR-1997; 97US-0818111.

PR 11-OCT-1996; 96US-0729622.

XX (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;

PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

DR N-PSDB; AAV44421.

XX WPI; 1998-251292/22.

XX N-PSDB; AAV44421.

PS Example 3; Page 193-194; 250pp; English.

CC This polypeptide is encoded by Mycobacterium tuberculosis antigen  
CC XP36 DNA (see AAV44421) that was isolated from a M. tuberculosis  
CC strain Erdman genomic DNA expression library using sera from  
CC patients having extrapulmonary tuberculosis. XP36 bears no  
CC similarity to known sequences. The invention relates to methods  
CC and compositions for diagnosing tuberculosis. It provides  
CC polypeptides (see AAW64291-64379) comprising an antigenic portion of a  
CC soluble M. tuberculosis antigen, or an immunogenic portion of a  
CC M. tuberculosis antigen, as well as DNA sequences encoding such  
CC polypeptides, recombinant expression vectors and transformed or  
CC transfected host cells. Also claimed are methods and diagnostic  
CC kits for detecting M. tuberculosis infection in a patient using  
CC these polypeptides, antibodies or oligonucleotide probes and  
CC primers.

SQ Sequence 196 AA;

Query Match 90.0%; Score 27; DB 19; Length 196;

Best Local Similarity 83.3%; Pred. No. 6.3e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GSIGS 6  
Db 190 gsvgs 195

## RESULT 35

AAV39167  
ID AAV39167 standard; Protein: 196 AA.

XX AAV39167;

DT 05-NOV-1999 (first entry)

DE M. tuberculosis antigen XP36 2nd predicted amino acid sequence.

XX Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;

KW immunotherapy; diagnosis; immunisation; vaccine; infection;

XX immune response; skin test.

OS Mycobacterium tuberculosis.

XX WO942076-A2.

PD 26-AUG-1999.  
 XX  
 PF 17-FEB-1999; 99MO-US03268.  
 XX  
 PR 05-MAY-1998; 98US-0072967.  
 PR 18-FEB-1998; 98US-0025197.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;  
 PI Lodes MJ, Reed SG, Skeiky YAM, Twardzik DR, Vedvick TS;  
 DR WPI; 1999-527409/44.  
 XX  
 PT New antigens from Mycobacterium tuberculosis useful in diagnostic  
 PT skin tests and protective or therapeutic vaccines or compositions  
 XX  
 PS Example 3; Page 177-178; 299pp; English.  
 XX  
 CC The present invention describes polypeptides comprising an immunogenic  
 CC part of a Mycobacterium tuberculosis antigen (Ag). Also described  
 CC are vaccines and fusion protein containing M. tuberculosis Ag's.  
 CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and  
 CC other polypeptides fragments, can be used in pharmaceutical compositions  
 CC or vaccines to generate a protective or therapeutic immune response to  
 CC M. tuberculosis and as reagents in skin tests for diagnosis of  
 CC tuberculosis. Ag can induce proliferation of, or cytokine secretion  
 CC by T, B or natural killer cells and/or macrophages in  
 CC tuberculosis-immune subjects. AA219249 to AA219460 and AAY39083 to  
 CC AAY39225 are used in the exemplification of the present invention.  
 XX  
 SQ Sequence 196 AA;

Query Match 90.0%; Score 27; DB 20; Length 196;  
 Best Local Similarity 83.3%; Pred. No. 6.3e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLSGS 6  
 ||:||||  
 Db 190 gsvvgs 195

RESULT 36  
 AAY39024  
 ID AAY39024 standard; Protein; 196 AA.  
 XX  
 AC AAY39024;  
 XX  
 DT 05-NOV-1999 (first entry)  
 XX  
 DE M. tuberculosis recombinant antigen protein XP36 #2.  
 XX  
 KW Antigen; diagnosis; detection; infection; antibody; immunisation;  
 KW vaccine; immunity.  
 XX  
 OS Mycobacterium tuberculosis.  
 XX  
 PN MO9942118-A2.  
 PN  
 PD 26-AUG-1999.  
 PD  
 PF 17-FEB-1999; 99MO-US03265.  
 PF  
 PR 05-MAY-1998; 98US-0072596.  
 PR 18-FEB-1998; 98US-0024753.  
 XX  
 PA (CORI-) CORIXA CORP.  
 PA  
 PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;  
 PI Lodes MJ, Reed SG, Skeiky YAM, Twardzik DR, Vedvick TS;  
 XX  
 DR WPI; 1999-527416/44.

DR N-PSDB; AA219119.  
 XX  
 PT New polypeptide comprising antigenic portions of M. tuberculosis  
 PT  
 XX  
 PS Example 3; Page 223; 323pp; English.  
 XX  
 CC This invention describes novel recombinant antigens and their encoding  
 CC nucleic acids derived from Mycobacterium tuberculosis. The novel  
 CC polypeptides are useful for detecting M. tuberculosis infection in a  
 CC biological sample by detecting antibodies which bind with the  
 CC polypeptides, and are useful as vaccines for immunizing against  
 CC M. tuberculosis infection. The new detection methods are needed as  
 CC current vaccination strategies do not provide 100% immunity.  
 XX  
 SQ Sequence 196 AA;

Query Match 90.0%; Score 27; DB 20; Length 196;  
 Best Local Similarity 83.3%; Pred. No. 6.3e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLSGS 6  
 ||:||||  
 Db 190 gsvvgs 195

RESULT 37  
 AAB56951  
 ID AAB56951 standard; Protein; 233 AA.  
 XX  
 AC AAB56951;  
 XX  
 DT 13-MAR-2001 (first entry)  
 XX  
 DE Human prostate cancer antigen protein sequence SEQ ID NO:1529.  
 XX  
 KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;  
 KW neuroprotective; cytosolic; cardioprotective; immunomodulatory; muscular;  
 KW vulnary; gastrointestinal; nephrotoxic; antineoplastic; gynecological;  
 KW antibacterial; gene therapy; neural; immune; reproductive; renal;  
 KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;  
 KW wound; infectious disease.  
 XX  
 OS Homo sapiens.  
 OS  
 PN MO20005174-A1.  
 PN  
 PD 21-SEP-2000.  
 PD  
 PE 08-MAR-2000; 2000MO-US05988.  
 PE  
 PR 12-MAR-1999; 99US-0124270.  
 PR  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PA (ROSE/) ROSEN C A.  
 XX  
 PI Rosen CA, Ruben SM;  
 PI  
 DR WPI; 2000-587513/55.  
 DR N-PSDB; AAF16154.  
 DR  
 XX  
 PT Prostate cancer associated gene sequences, referred to as prostate  
 PT cancer antigens, useful for treatment, prevention, and diagnosis of  
 PT disorders such as prostate cancer -  
 XX  
 PS Claim 11; Page 1970-1971; 2338pp; English.  
 XX  
 CC AAF15566 to AAF16505 encode the human prostate cancer associated  
 CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.  
 CC The prostate cancer antigens can have neuroprotective, cytosolic,  
 CC cardioprotective, immunomodulatory, muscular, vulnary, gastrointestinal,  
 CC nephrotoxic, antineoplastic, gynecological and antibacterial activities,  
 CC and can be used in gene therapy. The prostate cancer antigen



CC polynucleotides may be used for detection of prostate cancer, chromosome  
 CC identification, as chromosome markers, and for numerous other diagnostic  
 CC or research purposes. The prostate cancer antigens may be used to treat  
 CC disorders such as neural, immune, muscular, reproductive,  
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative  
 CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to  
 CC AAB57303 represent sequences used in the exemplification of the present  
 CC invention.

SO Sequence 233 AA:

Query Match 90.0%; Score 27; DB 21; Length 233;

Best Local Similarity 83.3%; Pred. No. 7.6e+02; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSIGGS 6  
 Db 102 galggs 107

RESULT 38

ID AAY06887 standard; Protein: 235 AA.

AC AAY06887;

DT 30-JUN-1999 (first entry)

DE HMNHJ20 polypeptide fragment derived from EST.

KW HMNHJ20 polypeptide; pim family; cancer; autoimmune disease; asthma;

KM rheumatoid arthritis; Alzheimer's disease; AIDS; stroke; gene therapy;

OS genetic linkage; chromosome localisation.

XX Homo sapiens.

XX EP911391-A2.

XX 28-APR-1999.

XX 19-OCT-1998; 98BP-0308550.

XX 27-OCT-1997; 97US-0123184.

XX 24-OCT-1997; 97US-0063245.

XX (SMIK ) SMITHKLINE BEECHAM CORP.

XX Albione EF, Kikly KK;

XX WPI; 1999-246411/21.

XX N-PSDB; AAX32973.

XX New clone, HMNHJ20, useful for diagnosing and treating cancer, AIDS

XX and autoimmune diseases

XX Claim 14; Page 15; 21pp; English.

XX The invention relates to a HMNHJ20 polypeptide, a member of the pim

XX family of polypeptides. Host cells transfected with a vector comprising

XX the HMNHJ20 nucleic acid are used for the recombinant expression of the

XX polypeptide. The HMNHJ20 nucleic acid may be used as probes or primers to

XX screen for genetic mutations in, or alterations in the expression of the

XX gene, in a patient compared to a healthy individual. This may lead to the

XX diagnosis of or identification of a predisposition to diseases such as

XX cancer, autoimmune disease, asthma, rheumatoid arthritis, Alzheimer's

CC alterations in gene expression; and may also be used to localize genes  
 CC on chromosomes. The present sequence represents a HMNHJ20 fragment  
 CC derived from EST.

SO Sequence 235 AA:

Query Match 90.0%; Score 27; DB 20; Length 235;

Best Local Similarity 83.3%; Pred. No. 7.6e+02; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSIGGS 6  
 Db 80 galgga 85

RESULT 39

ID AAM53007 standard; Protein: 246 AA.

AC AAM53007;

DT 03-AUG-1998 (first entry)

DE Mus musculus I-mfa protein.

KW I-mfa; inhibitor of MyoD family; treatment; diagnosis; myogenesis;

KM defects; abnormal development; disease; cleidocranial dysplasia;

XX CDD; Rhabdomyosarcoma; muscle tissue cancer.

XX Mus musculus.

XX W09808860-A1.

XX 05-MAR-1998.

XX 21-AUG-1997; 97MO-US14780.

XX 27-AUG-1996; 96US-0704931.

XX (HUTC-) HUTCHINSON CANCER RES CENT FRID.

XX Chen CMA, Groudine M, Kraut N, Weintrub H;

XX WPI; 1998-179377/16.

XX N-PSDB; AAV21282.

XX Inhibitor of MyoD family proteins - useful for, e.g., treatment and

XX diagnosis of defects in myogenesis responsible for abnormal

XX development

XX Disclosure: Pages 74-75; 92pp; English.

XX The sequence is that of murine I-mfa (inhibitor of MyoD family)

XX protein. Probes from the gene sequence can be used for determining the

XX presence of an I-mf protein or analogue, or for detecting I-mf

XX agonist activity in a test substance. The sequence can be used

XX to treat, model and diagnose defects in myogenesis responsible

XX for abnormal development and disease conditions in mammals,

XX including humans. These include the human disease cleidocranial

XX dysplasia (CDD), rhabdomyosarcomas and other cancers affecting

XX muscle tissue in mammals.

SO Sequence 246 AA:

Query Match 90.0%; Score 27; DB 19; Length 246;

Best Local Similarity 83.3%; Pred. No. 8e+02; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSIGGS 6  
 Db 120 galggs 125

## RESULT 40

AAW53008

ID AAW53008 standard; Protein: 251 AA.

XX

AAW53008;

AC

XX

03-AUG-1998 (first entry)

DT

XX

Mus musculus I-mfb protein.

DE

XX

I-mfb; Inhibitor of MyoD family; treatment; diagnosis; myogenesis; defects; abnormal development; disease; cleidocranial dysplasia;

KW

CD; rhabdomyosarcoma; muscle tissue cancer.

XX

OS

XX

Mus musculus.

PN

XX

W09808860-A1.

PD

XX

05-MAR-1998.

PE

XX

21-AUG-1997; 97WO-US14780.

PR

XX

27-AUG-1996; 96US-0704931.

PA

XX

(HUTC-) HUTCHINSON CANCER RES CENT FRED.

PI

XX

Chen CMA, Groudine M, Kraut N, Weintraub H;

DR

XX

WPI; 1998-179377/16.

DR

XX

N-PSDB; AAV21283.

PT

XX

Inhibitor of MyoD family proteins - useful for, e.g., treatment and diagnosis of defects in myogenesis responsible for abnormal development

PT

XX

Disclosure: Pages 77-78; 92pp; English.

PS

XX

The sequence is that of murine I-mfb (inhibitor of MyoD family) protein. Probes from the gene sequence can be used for determining the

CC

XX

presence of an I-mf protein or analogue, or for detecting I-mf

CC

XX

agonist activity in a test substance. The sequence can be used

CC

XX

to treat, model and diagnose defects in myogenesis responsible

CC

XX

for abnormal development and disease conditions in mammals,

CC

XX

including humans. These include the human disease cleidocranial

CC

XX

dysplasia (CD), rhabdomyosarcomas and other cancers affecting

CC

XX

muscle tissue in mammals.

XX

Sequence 251 AA:

SO

XX

Query Match 90.0%; Score 27; DB 19; Length 251;

OY

XX

Best Local Similarity 83.3%; Pred. No. 8.2e+02;

DB

XX

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

XX

1 GSLGGS 6

XX

120 galsgs 125

XX

Search completed: February 4, 2002, 08:03:05

XX

Job time: 169 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 4, 2002, 08:01:10 : Search time 12.9 Seconds

(without alignments)  
17.053 Million cell updates/sec

Title: us-09-642-660-11  
Sequence: 1 GSIIGS 6

Scoring table: BIOSIM62  
Gapop 10.0, Gapext 0.5

Searched: 100059 seqs, 36664827 residues  
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database : **SwissProt\_39.\***  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	30	100.0	444 1	YVBA_BACSU
2	30	100.0	452 1	PTCC_BACSU
3	30	100.0	993 1	HGBB_HAEIN
4	30	100.0	999 1	HGP2_HAEIN
5	30	100.0	999 1	HGPB_HAEIN
6	30	100.0	1013 1	HGBA_HAEIN
7	28	93.3	166 1	LSHB_CORTA
8	28	93.3	425 1	POKN_CORTA
9	28	93.3	572 1	STP_HAEIN
10	28	93.3	633 1	CZAD_BACTU
11	28	93.3	758 1	YKD4_YEAST
12	28	93.3	1402 1	ATCX_SCHRO
13	27	90.0	116 1	YIG7_YEAST
14	27	90.0	149 1	YMH3_THIRE
15	27	90.0	207 1	HIS7_AZORR
16	27	90.0	281 1	STRE_STRGR
17	27	90.0	290 1	HO3_RAT
18	27	90.0	312 1	YMH2_MYCTU
19	27	90.0	315 1	HO2_MOUSE
20	27	90.0	315 1	HO2_RAT
21	27	90.0	345 1	ESTR_STRSC
22	27	90.0	378 1	NAGA_VIBCH
23	27	90.0	396 1	PCIR_TOBAC
24	27	90.0	404 1	RAGE_HUMAN
25	27	90.0	404 1	YAXM_HITSN
26	27	90.0	448 1	YCDG_ECOLI
27	27	90.0	458 1	MSRE_MOUSE
28	27	90.0	466 1	CHLN_SYNP7
29	27	90.0	482 1	K2C8_RAT
30	27	90.0	495 1	CD5_HUMAN
31	27	90.0	501 1	HDAC_ARATH
32	27	90.0	508 1	USP_DROME
33	27	90.0	580 1	P69_MYCHR

34	27	90.0	643 1	TRPA_EPTST	090501 eptatretus
35	27	90.0	673 1	FXO3_HUMAN	043524 homo sapien
36	27	90.0	726 1	FARA_VIBAN	P11461 vibrio angu
37	27	90.0	773 1	PIGR_RABIT	P01832 oryctolagus
38	27	90.0	900 1	FOX2_YEAST	002207 saccharomyc
39	27	90.0	910 1	AUX1_BOVIN	027974 bos taurus
40	27	90.0	972 1	HGBA_HAEIN	047952 haemophilus
41	27	90.0	972 1	HGBB_HAEIN	047957 haemophilus
42	27	90.0	985 1	DPOL_HSV1	P28859 ictaluriid h
43	27	90.0	999 1	HGP4_HAEIN	057408 haemophilus
44	27	90.0	1046 1	HUVA_HAEIN	048153 haemophilus
45	27	90.0	1063 1	HGPI_HAEIN	P44795 haemophilus
46	27	90.0	1067 1	HGPC_HAEIN	094442 haemophilus
47	27	90.0	1067 1	HGBB_HAEIN	094221 haemophilus
48	27	90.0	1077 1	HGPA_HAEIN	094221 haemophilus
49	27	90.0	1084 1	HGP3_HAEIN	P44836 haemophilus
50	27	90.0	1305 1	GAK_RAT	P97874 rattus norv
51	27	90.0	1584 1	KYK1_DICDI	P18160 dictyostell
52	27	90.0	1645 1	OMP_B_RICRY	P66989 r outer mem
53	27	90.0	1654 1	OMP_B_RICRI	053047 r outer mem
54	27	90.0	1655 1	OMP_B_RICCN	093K43 r outer mem
55	27	90.0	1656 1	OMP_B_RICJA	P06653 r outer mem
56	27	90.0	3726 1	TRX_DROME	P20659 drosophila
57	26	86.7	176 1	PACA_HUMAN	P18509 h pituitary
58	26	86.7	194 1	IM22_HUMAN	094584 homo sapien
59	26	86.7	209 1	COBC_SCYCA	P81130 scyllorhinu
60	26	86.7	210 1	CRB3_BOVIN	P26998 homo sapien
61	26	86.7	211 1	CRB3_HUMAN	P19141 bos taurus
62	26	86.7	211 1	CRB3_MOUSE	091149 mus musculu
63	26	86.7	211 1	CRB3_RAT	P05524 rattus norv
64	26	86.7	211 1	YCGA_ECOLI	P11667 escherichia
65	26	86.7	212 1	OMPW_ECOLI	P12364 escherichia
66	26	86.7	222 1	Y561_DEIRA	094722 delnoccocus
67	26	86.7	233 1	Y6G1_ECOLI	P46143 escherichia
68	26	86.7	273 1	LECG_ARAHY	P02872 arachis hyp
69	26	86.7	280 1	BPHB_RHOGO	P47230. r cis-2,3-d
70	26	86.7	298 1	YCSB_CAEEL	022111 caenorhabdi
71	26	86.7	317 1	Y274_PYRHO	056012 pyrococcus
72	26	86.7	324 1	Y998_PYRAB	094225 pyrococcus
73	26	86.7	333 1	YRP5_IRVB	P18309 chilo iride
74	26	86.7	341 1	NU2M_ANOQU	P34848 anopheles g
75	26	86.7	341 1	NU2M_ANOQU	P33503 anopheles g
76	26	86.7	346 1	PUR5_BACSU	P12043 bacillus su
77	26	86.7	359 1	LEU3_KLUMA	P41766 kluyveromyc
78	26	86.7	362 1	LEU3_KLUMA	P23390 kluyveromyc
79	26	86.7	364 1	LEU3_KLUMA	P41773 saccharomyc
80	26	86.7	368 1	LEU3_YEAST	094230 kluyveromyc
81	26	86.7	378 1	YEHF_ECOLI	P33352 escherichia
82	26	86.7	379 1	METX_NEIMA	094295 neisseria m
83	26	86.7	379 1	METX_NEIMA	094295 neisseria m
84	26	86.7	379 1	METX_PSENE	P57714 pseudomonas
85	26	86.7	380 1	METX_PSENE	094295 neisseria m
86	26	86.7	380 1	KCRB_CHICK	P05122 gallus gall
87	26	86.7	445 1	YDDE_ECOLI	P39266 escherichia
88	26	86.7	448 1	WTL_RAT	P49955 rattus norv
89	26	86.7	449 1	WTL_HUMAN	P19544 homo sapien
90	26	86.7	449 1	WTL_MOUSE	P22561 mus musculu
91	26	86.7	454 1	YAJR_ECOLI	062651 sus scrofa
92	26	86.7	454 1	YAJR_ECOLI	P77726 escherichia
93	26	86.7	456 1	SECY_MERTH	061334 methanobact
94	26	86.7	457 1	CGPB_FUSO	000858 fusarium so
95	26	86.7	482 1	SR54_SYNY3	P74214 synechocyst
96	26	86.7	495 1	POLG_DEN21	P14339 dengue viru
97	26	86.7	495 1	POLG_DEN23	P14339 dengue viru
98	26	86.7	511 1	YE08_YEAST	P40051 saccharomyc
99	26	86.7	511 1	CHX2_MOUSE	P30658 mus musculu
100	26	86.7	526 1	KICJ_BOVIN	P06394 bos taurus

## ALIGNMENTS

RESULT 1

YBWA\_BACSU STANDARD: PRT: 444 AA.

AC P3984; 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE PUTATIVE PTS SYSTEM IIB COMPONENT YBWA (EC 2.7.1.69).

GN YBWA OR IPA-16D.

OS Bacillus subtilis.

OC Bacteria: Firmicutes: Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group: Bacillus.

OX NCBI\_TaxID=1423;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX MEDLINE=95020537; PubMed=7934828;

RA Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,

RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer I.,

RA Prescan E., Santana M., Schneider E., Schweizer J., Vertes A.,

RA Rapoport G., Danchin A.;

RT "Bacillus subtilis genome project: cloning and sequencing of the 97

RT kb region from 325 degrees to 333 degrees.";

RL MOJ. Microbiol. 10:321-384(1993).

CC -1- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOTRANSFERASE + SUGAR -

CC PROTEIN HISTIDINE + SUGAR PHOSPHATE.

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).

CC -1- SIMILARITY: CONTAINS A PTS IIB DOMAIN.

CC -1- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: X73124; CAAS1572.1; -;

DR EMBL: Z99123; CAB15865.1; -;

DR PIR: S39671; S39671.

DR Subtilist; BG10562; YBWA.

DR InterPro: IPR003352; PTS\_EIIC.

DR Pfam: PF02378; PTS\_EIIC; 1.

KW Hypothetical protein; Phosphotransferase system; Sugar transport;

KW Transferrase; Phosphorylation; Transmembrane; Complete proteome.

CC CC

FT DOMAIN 1 444

FT TRANSMEM 31 51 POTENTIAL.

FT TRANSMEM 72 92 POTENTIAL.

FT TRANSMEM 104 124 POTENTIAL.

FT TRANSMEM 138 158 POTENTIAL.

FT TRANSMEM 187 207 POTENTIAL.

FT TRANSMEM 223 243 POTENTIAL.

FT TRANSMEM 246 266 POTENTIAL.

FT TRANSMEM 291 311 POTENTIAL.

FT TRANSMEM 338 358 POTENTIAL.

FT TRANSMEM 379 399 POTENTIAL.

FT TRANSMEM 402 422 POTENTIAL.

SO SEQUENCE 444 AA; 47647 MM; D3964B3795274C7D CRC64;

Query Match 100.0%; Score 30; DB 1; Length 444;

Best Local Similarity 100.0%; Pred. No. 61;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLGGS 6

DB 232 GSLGGS 237

RESULT 2

PTCC\_BACSU STANDARD: PRT: 452 AA.

AC P46317; 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE PTS SYSTEM, CELLOBIOSE-SPECIFIC IIC COMPONENT (EIIC-CEL) (CELLOBIOSE-

DE PERMEASE IIC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, C COMPONENT).

GN CELB OR LICC.

OS Bacillus subtilis.

OC Bacteria: Firmicutes: Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group: Bacillus.

OX NCBI\_TaxID=1423;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX MEDLINE=97144536; PubMed=8990303;

RA Tobisch S., Glaser P., Krueger S., Hecker M.;

RT "Identification and characterization of a new beta-glucoside

RT utilization system in Bacillus subtilis.";

RL J. Bacteriol. 179:496-506(1997).

RL [2]

RP SEQUENCE FROM N.A.

RC STRAIN=BGSC 1A1;

RX MEDLINE=97124196; PubMed=8969509;

RA Yoshida K., Shindo K., Sano H., Seki S., Fujimura M., Yanai N.,

RA Miwa Y., Fujita Y.;

RT "Sequencing of a 65 kb region of the Bacillus subtilis genome

RT containing the iic and cel loci, and creation of a 177 kb contig

RT covering the gnt-sacxy region.";

RL Microbiology 142:3113-3123(1996).

CC -1- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOTRANSFERASE-DEPENDENT

CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS). A MAJOR CARBOHYDRATE ACTIVE

CC -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE

CC AND THE TRANSMEMBRANE CHANNEL. THE IIA DOMAIN CONTAINS THE PRIMARY

CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS

CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO

CC THE SUGAR.

CC CC

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).

CC -1- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: Z49992; CAA90286.1; -;

DR EMBL: D83026; BAA11744.1; -;

DR EMBL: Z99123; CAB15884.1; -;

DR Subtilist; BG1348; CELB.

DR InterPro: IPR003352; PTS\_EIIC.

DR Pfam: PF02378; PTS\_EIIC; 1.

KW Phosphotransferase system; Sugar transport; Transmembrane;

KW Complete proteome.

CC CC

FT TRANSMEM 31 51 POTENTIAL.

FT TRANSMEM 72 92 POTENTIAL.

FT TRANSMEM 104 124 POTENTIAL.

FT TRANSMEM 138 158 POTENTIAL.

FT TRANSMEM 187 207 POTENTIAL.

FT TRANSMEM 218 238 POTENTIAL.

FT TRANSMEM 246 266 POTENTIAL.

FT TRANSMEM 292 312 POTENTIAL.

FT TRANSMEM 339 359 POTENTIAL.

FT TRANSMEM 379 399 POTENTIAL.

FT TRANSMEM 402 422 POTENTIAL.

SO SEQUENCE 452 AA; 48534 MM; 396AE5531486C429 CRC64;

Query Match 100.0%; Score 30; DB 1; Length 452;

Best Local Similarity 100.0%; Pred. No. 62;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLGGS 6  
 DB 232 GSLGGS 237

RESULT 3  
 HGBC\_HAEIN STANDARD; PRT; 993 AA.  
 AC Q9KIV0;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DE 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE HEMOGLOBIN BINDING PROTEIN C PRECURSOR.  
 GN HGBC.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus.  
 ON NCBI\_TaxID=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-NTHI N182; PubMed=10858226;  
 RX MEDLINE=20316037; Cope L.D., Hrkal Z., Hansen E.J.;  
 RA Cope L.D., Hrkal Z., Hansen E.J.;  
 RT Detection of phase variation in expression of proteins involved in  
 hemoglobin and hemoglobin-haptoglobin binding by nontypeable  
 Haemophilus influenzae.\*;  
 RL Infect. Immun. 68:4092-4101(2000).  
 CC -1- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OF THE HUMAN HOST AND  
 IS REQUIRED FOR HEME UPTAKE.  
 CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.  
 CC EXPRESSION ASSOCIATED WITH ALTERATION IN THE LENGTH OF THE CCA  
 REPEAT REGION. THIS MECHANISM IS CALLED SLIPPED-STRAND MISPAIRING.  
 CC ADDITION OR LOSS OF CCA REPEAT UNITS WOULD CHANGE THE READING  
 FRAME AND RESULT IN INTRODUCTION OF STOP CODONS DOWNSTREAM OF THE  
 REPEAT REGION. THIS MAY BE A MECHANISM OF REGULATION AND A  
 WAY TO AVOID THE IMMUNOLOGICAL RESPONSE OF THE HOST.  
 CC -1- SIMILARITY: BELONGS TO THE TONB-DEPENDENT RECEPTOR PROTEIN FAMILY;  
 HEMOGLOBIN/HAPTOGLOBIN BINDING PROTEIN SUBFAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 DR EMBL: AF221060; AAF80178.1; -  
 DR InterPro: IPR000531; Tonb\_boxC.  
 DR Pfam: PF00593; Tonb\_boxC.1.  
 DR PROSITE: PS00430; TONB\_DEPENDENT\_REC\_1; FALSE\_NEG.  
 DR PROSITE: PS01156; TONB\_DEPENDENT\_REC\_2; 1.  
 KW Outer membrane; Transport; Tonb box; Multigene family; Signal;  
 KM Receptor; Repeat.  
 FT SIGNAL 1 24  
 FT CHAIN 25 993  
 FT DOMAIN 26 37  
 FT REPEAT 26 29  
 FT REPEAT 30 33  
 FT REPEAT 34 37  
 FT SITE 47 54  
 FT SITE 976 993  
 SQ SEQUENCE 993 AA; 113616 MW; A51BF3B2C641612 CRC64;

Query Match 100.0%; Score 30; DB 1; Length 993;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLGGS 6  
 DB 174 GSLGGS 179

RESULT 4  
 HGP2\_HAEIN STANDARD; PRT; 999 AA.  
 AC P44809;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE PROBABLE HEMOGLOBIN AND HEMOGLOBIN-HAPTOGLOBIN BINDING PROTEIN 2  
 DE PRECURSOR.  
 GN H10661.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus.  
 ON NCBI\_TaxID=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-RD / KW20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 Kierlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghegan N.S.M.,  
 Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 Venter J.C.;  
 RT Whole-genome random sequencing and assembly of Haemophilus  
 RT influenzae Rd.\*;  
 RL Science 269:496-512(1995).  
 RN [2]  
 RP REVISIONS.  
 RA White O., Clayton R.A., Kierlavage A.R., Fleischmann R.D.,  
 Peterson J., Hickey E., Dodson R., Gwin M.;  
 RA Submitted (MAY-1998) to the EMBL/Genbank/DBD databases.  
 RN [3]  
 RP IDENTIFICATION BY MASS SPECTROMETRY.  
 RX MEDLINE=20137488; PubMed=10675023;  
 RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,  
 Gray C., Fountoulakis M.;  
 RA Two-dimensional map of the proteome of Haemophilus influenzae.\*;  
 RL Electrophoresis 21:411-429(2000).  
 RN [4]  
 RP CONCEPTUAL TRANSLATION.  
 RA Couderc E.;  
 RL Unpublished observations (AUG-2001).  
 CC -1- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OR THE  
 HEMOGLOBIN/HAPTOGLOBIN COMPLEX OF THE HUMAN HOST AND IS REQUIRED  
 FOR HEME UPTAKE (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).  
 CC EXPRESSION ASSOCIATED WITH ALTERATION IN THE LENGTH OF THE CCA  
 REPEAT REGION. THIS MECHANISM IS CALLED SLIPPED-STRAND MISPAIRING.  
 CC ADDITION OR LOSS OF CCA REPEAT UNITS WOULD CHANGE THE READING  
 FRAME AND RESULT IN INTRODUCTION OF STOP CODONS DOWNSTREAM OF THE  
 REPEAT REGION. THIS MAY BE A MECHANISM OF REGULATION AND A WAY TO  
 AVOID THE IMMUNOLOGICAL RESPONSE OF THE HOST (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE TONB-DEPENDENT RECEPTOR PROTEIN FAMILY;  
 HEMOGLOBIN/HAPTOGLOBIN BINDING PROTEIN SUBFAMILY.  
 CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION: TWO FRAMESHIFTS WERE  
 INTRODUCED IN THE REPEATS REGION AND IN THE LEADER PEPTIDE TO  
 MAXIMIZE THE SIMILARITY WITH OTHER ORTHOLOGS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC

DR EMBL: U32749; AAC22319.1; ALT\_SEQ.  
DR TIGR: H10661; -  
DR InterPro: IPR000531; TonB\_boxC.  
DR Pfam: PF00593; TonB\_boxC; 1.  
DR PROSITE: PS00430; TONB\_DEPENDENT\_REC\_1; FALSE\_NEG.  
DR PROSITE: PS01156; TONB\_DEPENDENT\_REC\_2; 1.  
KW Outer membrane; Transport; TonB box; Multigene family; Signal;  
KW Receptor; Repeat; Complete proteome.  
FT SIGNAL 1 24 POTENTIAL.  
FT CHAIN 25 999 PROBABLE HEMOGLOBIN AND HEMOGLOBIN-  
HAPTOGLOBIN BINDING PROTEIN 2.  
FT DOMAIN 26 53 1.  
FT REPEAT 26 29 7 X 4 AA TANDEM REPEATS OF Q-P-T-N.  
FT REPEAT 30 33 2.  
FT REPEAT 34 37 3.  
FT REPEAT 38 41 4.  
FT REPEAT 42 45 5.  
FT REPEAT 46 49 6.  
FT REPEAT 50 53 7.  
FT SITE 63 70 TONB BOX.  
FT SITE 982 999 TONB C-TERMINAL BOX.  
SQ SEQUENCE 999 AA; 114690 MW; 1A17AAB220092B7D CRC64;  
  
Query Match 100.0%; Score 30; DB 1; Length 999;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GSLSGS 6  
DB 193 GSLSGS 195  
IIIIII  
190 GSLSGS 195  
  
RESULT 5  
HGPA\_HAEIN STANDARD: PRT; 999 AA.  
ID HGPA\_HAEIN  
AC 087296;  
DT 20-AUG-2001 (Rel. 40, Created)  
DT 20-AUG-2001 (Rel. 40, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE HEMOGLOBIN AND HEMOGLOBIN-HAPTOGLOBIN BINDING PROTEIN B PRECURSOR.  
GN HGPA.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OC NCBI\_Taxid=727;  
OX NCBI\_Taxid=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H1689 / serotype B;  
RX MEDLINE=98427137; PubMed=9746572;  
RA Ren Z., Jin H., Morton D.J., Stull T.L.;  
RT "hgpa", a gene encoding a second Haemophilus influenzae hemoglobin- and  
RT hemoglobin-haptoglobin-binding protein.\*;  
RL Infect. Immun. 66:4733-4741(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ElA / serotype B;  
RA Morton D.J., Stull T.L.;  
RT "Conservation of hemoglobin/hemoglobin-haptoglobin binding proteins in  
RT Haemophilus influenzae.";  
RL Submitted (APR-2000) to the EMBL/Genbank/DBJ databases.  
CC -1- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OR THE  
CC HEMOGLOBIN/HAPTOGLOBIN COMPLEX OF THE HUMAN HOST AND IS REQUIRED  
CC FOR HEME UPTAKE.  
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.  
CC -1- MISCELLANEOUS: THIS PROTEIN IS SUBJECT TO PHASE-VARIABLE  
CC EXPRESSION ASSOCIATED WITH ALTERATION IN THE LENGTH OF THE CCAA  
CC REPEAT REGION. THIS MECHANISM IS CALLED SLIPPED-STRAND MISPAIRING.  
CC ADDITION OR LOSS OF CCAA REPEAT UNITS WOULD CHANGE THE READING  
CC FRAME AND RESULT IN INTRODUCTION OF STOP CODONS DOWNSTREAM OF THE  
CC REPEAT REGION. THIS MAY BE A MECHANISM OF REGULATION AND A  
CC WAY TO AVOID THE IMMUNOLOGICAL RESPONSE OF THE HOST.  
CC -1- SIMILARITY: BELONGS TO THE TONB-DEPENDENT RECEPTOR PROTEIN FAMILY;  
CC

CC HEMOGLOBIN/HAPTOGLOBIN BINDING PROTEIN SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: AF022910; AAC60790.1; -  
DR EMBL: AF259266; AAK51630.1; -  
DR InterPro: IPR000531; TonB\_boxC.  
DR Pfam: PF00593; TonB\_boxC; 1.  
DR PROSITE: PS00430; TONB\_DEPENDENT\_REC\_1; FALSE\_NEG.  
DR PROSITE: PS01156; TONB\_DEPENDENT\_REC\_2; 1.  
KW Outer membrane; Transport; TonB box; Multigene family; Signal;  
KW Receptor; Repeat.  
FT SIGNAL 1 24 POTENTIAL.  
FT CHAIN 25 999 HEMOGLOBIN AND HEMOGLOBIN-HAPTOGLOBIN  
BINDING PROTEIN B.  
FT DOMAIN 26 57 8 X 4 AA TANDEM REPEATS OF P-T-N-Q.  
FT REPEAT 26 29 1.  
FT REPEAT 30 33 2.  
FT REPEAT 34 37 3.  
FT REPEAT 38 41 4.  
FT REPEAT 42 45 5.  
FT REPEAT 46 49 6.  
FT REPEAT 50 53 7.  
FT REPEAT 54 57 8.  
FT SITE 66 73 TONB BOX.  
FT SITE 982 999 TONB C-TERMINAL BOX.  
FT VARIANT 46 57 MISSING (IN STRAIN ElA).  
FT VARIANT 85 57 I -> V (IN STRAIN ElA).  
FT VARIANT 101 101 D -> N (IN STRAIN ElA).  
FT VARIANT 108 108 T -> K (IN STRAIN ElA).  
FT VARIANT 121 122 LA -> FG (IN STRAIN ElA).  
FT VARIANT 151 151 T -> S (IN STRAIN ElA).  
FT VARIANT 304 304 G -> A (IN STRAIN ElA).  
FT VARIANT 560 560 N -> D (IN STRAIN ElA).  
FT VARIANT 633 633 L -> F (IN STRAIN ElA).  
FT VARIANT 638 638 R -> H (IN STRAIN ElA).  
FT VARIANT 689 689 K -> E (IN STRAIN ElA).  
FT VARIANT 755 755 A -> T (IN STRAIN ElA).  
FT VARIANT 782 782 S -> G (IN STRAIN ElA).  
SQ SEQUENCE 999 AA; 114435 MW; 58F631FA5D2685B0 CRC64;  
  
Query Match 100.0%; Score 30; DB 1; Length 999;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GSLSGS 6  
DB 193 GSLSGS 198  
IIIIII  
193 GSLSGS 198  
  
RESULT 6  
HGPA\_HAEIN STANDARD: PRT; 1013 AA.  
ID HGPA\_HAEIN  
AC Q9KIV2;  
DT 20-AUG-2001 (Rel. 40, Created)  
DT 20-AUG-2001 (Rel. 40, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE HEMOGLOBIN BINDING PROTEIN A PRECURSOR.  
GN HGPA.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OC NCBI\_Taxid=727;  
OX NCBI\_Taxid=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NTHI N182;  
CC

```

RX MEDLINE-20316037; PubMed-10858226;
RA Cope L.D., Hrkai Z., Hansen E.J.;
RT "Detection of phase variation in expression of proteins involved in
RT hemoglobin and hemoglobin-haptoglobin binding by non-typeable
RT Hemophilus influenzae."
RL Infect. Immun. 68:4092-4101(2000).
CC -1- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OF THE HUMAN HOST AND
CC IS REQUIRED FOR HEME UPTAKE.
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC MISCELLANEOUS: THIS PROTEIN IS SUBJECT TO PHASE-VARIABLE
CC EXPRESSION ASSOCIATED WITH ALTERATION IN THE LENGTH OF THE CCA
CC REPEAT REGION. THIS MECHANISM IS CALLED SLIPPED-STRAND MISPAIRING.
CC ADDITION OR LOSS OF CCA REPEAT UNITS WOULD CHANGE THE READING
CC FRAME AND RESULT IN INTRODUCTION OF STOP CODONS DOWNSTREAM OF THE
CC REPEAT REGION. THIS MAY BE A MECHANISM OF REGULATION AND A
CC WAY TO AVOID THE IMMUNOLOGICAL RESPONSE OF THE HOST.
CC -1- SIMILARITY: BELONGS TO THE TONB-DEPENDENT RECEPTOR PROTEIN FAMILY;
CC HEMOGLOBIN/HAPTOGLOBIN BINDING PROTEIN SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR EMBL: AF21059; AAF80176.1; -
DR InterPro: IPR000531; TonB_boxc.
DR Pfam: PF00593; TonB_boxc; 1.
DR PROSITE: PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
DR PROSITE: PS01156; TONB_DEPENDENT_REC_2; 1.
KW Outer membrane; Transport; TonB box; Multigene family; Signal;
KW Receptor; Repeat.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 1013 HEMOGLOBIN BINDING PROTEIN A.
FT DOMAIN 26 57 8 x 4 AA TANDEM REPEATS OF Q-P-T-N.
FT REPEAT 26 57 1.
FT REPEAT 26 57 2.
FT REPEAT 30 33 3.
FT REPEAT 34 37 4.
FT REPEAT 38 41 5.
FT REPEAT 42 45 6.
FT REPEAT 46 49 7.
FT REPEAT 50 53 8.
FT REPEAT 54 57 7.
FT REPEAT 54 57 8.
FT SITE 67 74 TONB BOX.
FT SITE 996 1013 TONB C-TERMINAL BOX.
SQ SEQUENCE 1013 AA; 116260 MW; 769964335AED5C1 CRC64;

```

```

Query Match 100.0%; Score 30; DB 1; Length 1013;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 GSLGGS 6
DB 194 GSLGGS 199

```

```

RESULT 7
LSH_COTJA STANDARD; PRT; 166 AA.
AC P45657;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE LUTROPIN BETA CHAIN PRECURSOR (LUTEINIZING HORMONE BETA SUBUNIT) (LSH-
DE BETA) (LSH-B) (LH-B).
GN LSH.
OS Coturnix coturnix japonica (Japanese quail).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Coturnix.

```

```

OX NCBI_TaxID-93934;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-94252550; PubMed-7515015;
RA Ando H., Ishii S.;
RT "Molecular cloning of complementary deoxyribonucleic acids for the
RT pituitary glycoprotein hormone alpha-subunit and luteinizing hormone
RT beta-subunit precursor molecules of Japanese quail (Coturnix coturnix
RT japonica)."
RL Gen. Comp. Endocrinol. 93:357-368(1994).
CC -1- FUNCTION: PROMOTES SPERMATOGENESIS AND OVULATION BY STIMULATING
CC THE TESTES AND OVARIES TO SYNTHESIZE STEROIDS.
CC -1- SUBUNIT: HETERODIMER OF A COMMON ALPHA CHAIN AND A UNIQUE BETA
CC CHAIN WHICH CONFERS BIOLOGICAL SPECIFICITY TO THYROTROPIN,
CC LUTROPIN, FOLLITROPIN AND GONADOTROPIN.
CC -1- SIMILARITY: BELONGS TO THE GLYCOPROTEIN HORMONES BETA CHAIN
CC FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR EMBL: S70834; AB330867.1; -
DR HSP: P01233; IHRP.
DR InterPro: IPR000359; Cys_knot.
DR InterPro: IPR002400; GF_CysKnot.
DR InterPro: IPR001545; Glyco_hormone_beta.
DR Pfam: PF00007; Cys_knot; 1.
DR PRINTS: PR00438; GRCYSKNOT.
DR SMART: SM00068; GHb; 1.
DR PROSITE: PS00261; GLYCO_HORMONE_BETA_1; 1.
DR PROSITE: PS00689; GLYCO_HORMONE_BETA_2; 1.
KW Hormone; Signal; Glycoprotein.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 166 LUTROPIN BETA CHAIN.
FT CARBOHYD 60 60 N-LINKED (GLCNAc... ) (POTENTIAL).
FT DISULFID 56 104 BY SIMILARITY.
FT DISULFID 70 119 BY SIMILARITY.
FT DISULFID 73 157 BY SIMILARITY.
FT DISULFID 81 135 BY SIMILARITY.
FT DISULFID 85 137 BY SIMILARITY.
FT DISULFID 140 147 BY SIMILARITY.
SQ SEQUENCE 166 AA; 17030 MW; 6BF293BEC3C5FACT CRC64;

```

```

Query Match 93.3%; Score 28; DB 1; Length 166;
Best Local Similarity 83.3%; Pred. No. 58;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 1 GSLGGS 6
DB 46 GSMGGS 51

```

```

RESULT 8
POXN_DROME STANDARD; PRT; 425 AA.
AC P23758;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PAIRED BOX POX-NEURO PROTEIN (PAIRED BOX NEURONAL PROTEIN).
DE POX-N.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID-7227;
RN [1]

```

RP SEQUENCE FROM N.A.  
 RX MEDLINE-92208941; PubMed-1348214;  
 RA Dambly-Chaudiere C., Jamet E., Burri M., Bopp D., Basler K.,  
 RT Hafen E., Dumont N., Spielmann P., Ghyzen A., Noll M.;  
 RL "The paired box gene *pox* encodes a determinant of poly-innervated  
 sense organs in *Drosophila*.";  
 RN Cell 69:159-172(1992).  
 [2]  
 RP SEQUENCE OF 1-35 FROM N.A.  
 RX MEDLINE-90059940; PubMed-2573516;  
 RA Bopp D., Jamet E., Baumgartner S., Burri M., Noll M.;  
 RT "Isolation of two tissue-specific *Drosophila* paired box genes, *pox*  
 and *meso* and *fox* neuro.";  
 RN EMBO J. 8:3447-3457(1989).  
 [3]  
 RP FUNCTION.  
 RX MEDLINE-94121914; PubMed-8292359;  
 RA Nitebohm E., Usui A., Therianos S., Kimura K., Dambly-Chaudiere C.,  
 RT Ghyzen A.;  
 RT "The gene *pox* controls different steps of the formation of  
 chemosensory organs in *Drosophila*.";  
 RL Neuron 12:25-34(1994).  
 [4]  
 RP FUNCTION.  
 RX MEDLINE-97326936; PubMed-9183748;  
 RA Awasaki T., Kimura K.;  
 RT "Pox-neuro is required for development of chemosensory bristles in  
*Drosophila*.";  
 RL J. Neurobiol. 32:707-721(1997).  
 CC -1- FUNCTION: TRANSCRIPTIONAL REGULATOR THAT SPECIFIES POLY-INNERVATED  
 CC ORGANS (CHEMOSENSORY BRISTLE). ALSO CONTROLS THE NUMBER OF  
 CC NEURONS.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- TISSUE SPECIFICITY: CENTRAL AND PERIPHERAL NERVOUS SYSTEM.  
 CC -1- SIMILARITY: CONTAINS A PAIRED BOX DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M86927; AAA28832.1; -;  
 DR PIR; S06951; S06951.  
 DR HSP; P06601; 1PDN.  
 DR FLYBASE; FBgn0003130; Pox-n.  
 DR InterPro: IPR001523; Paired\_Box.  
 DR Pfam; PF00292; PAX; 1.  
 DR PRINTS; PR00027; PAIREDBOX.  
 DR SMART; SM00351; PAX; 1.  
 DR PROSITE; PS00034; PAIRED\_BOX; 1.  
 DR KW Paired box; Developmental protein; Nuclear protein;  
 KW Transcription regulation; Neurogenesis; Differentiation.  
 FT DOMAIN 5 130 PAIRED\_BOX.  
 FT DOMAIN 146 157 POLY-ALA.  
 FT DOMAIN 173 176 POLY-PRO.  
 FT DOMAIN 281 287 POLY-ALA.  
 SQ SEQUENCE 425 AA; 44387 MW; EC76C6494A06DC15 CRC64;

Query Match 93.3%; Score 28; DB 1; Length 425;  
 Best Local Similarity 83.3%; Pred. No. 1.5e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSIGS 6  
 Db 70 GSIGS 75

RESULT 9

SYP\_HAEIN  
 ID\_SYP\_HAEIN STANDARD; PRT; 572 AA.  
 AC P43830;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE PROLYL-TRNA SYNTHETASE (EC 6.1.1.15) (PROLINE-TRNA LIGASE) (PRORS).  
 GN PROS OR H10729.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales;  
 OC Haemophilus.  
 OC NCBI\_TaxID:127;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-RD / KW20 / ATCC 51907;  
 RX MEDLINE-95350630; PubMed-7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of *Haemophilus*  
 RT *influenzae* Rd.";  
 RL Science 269:496-512(1995).  
 CC -1- CATALYTIC ACTIVITY: ATP + L-PROLINE + TRNA(PRO) = AMP +  
 CC PYROPHOSPHATE + L-PROLYL-TRNA(PRO).  
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U32756; AAC22386.1; -;  
 DR TIGR; H10729; -;  
 DR InterPro: IPR002106; AA\_TRNA\_Ligase\_II.  
 DR InterPro: IPR002314; tRNA-synt\_2b.  
 DR InterPro: IPR002316; tRNA-synt\_pro.  
 DR Pfam; PF00587; tRNA-synt\_2b; 1.  
 DR PRINTS; PRO1046; TRNAsyntPRO.  
 DR PROSITE; PS00179; AA\_TRNA\_LIGASE\_II\_1; 1.  
 DR PROSITE; PS00339; AA\_TRNA\_LIGASE\_II\_2; FALSE\_NEG.  
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
 KW Complete proteome.  
 SQ SEQUENCE 572 AA; 63975 MW; 5BFOF81162368583 CRC64;

Query Match 93.3%; Score 28; DB 1; Length 572;  
 Best Local Similarity 83.3%; Pred. No. 2e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSIGS 6  
 Db 200 GSIGS 205

RESULT 10  
 C2AD\_BACTU STANDARD; PRT; 633 AA.  
 ID C2AD\_BACTU  
 AC O9RMG3;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE PESTICIDIAL CRISTAL PROTEIN CRIZAD (INSECTICIDAL DELTA-ENDOTOXIN



```

DE CRYIA(D)) (CRYSTALLINE ENTOMOCIDAL PROTOXIN) (71 KDA CRYSTAL PROTEIN).
GN CRY2AD OR CRYIA(D) OR CRY2.
OS Bacillus thuringiensis.
OC Bacillus: Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1428;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BR30;
RA Chou S.-K., Shin B.-S., Park S.-H.;
RT "Nucleotide sequence of a new Bacillus thuringiensis cry2-type gene.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PROMOTES COLLOIDSMOTIC LYSIS BY BINDING TO THE MIDGUT
CC EPITHELIAL CELLS OF INSECTS.
CC -1- DEVELOPMENTAL STAGE: THE CRYSTAL PROTEIN IS PRODUCED DURING
CC SPOULATION AND IS ACCUMULATED BOTH AS AN INCLUSION AND AS PART
CC OF THE SPORE COAT.
CC -1- MISCELLANEOUS: TOXIC SEGMENT OF THE PROTEIN IS LOCATED IN THE
CC N-TERMINUS.
CC -1- SIMILARITY: BELONGS TO THE DELTA ENDOTOXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF200816; AAF09583.1; -
DR InterPro: IPR001178; Endotoxin.
DR Pfam: PF00555; endotoxin; 1.
DR Toxin; Sporulation.
KW TOXIN; Sporulation.
SQ SEQUENCE 633 AA; 70752 MW; 2A582067131B39CB CRC64;

Query Match
Best Local Similarity 93.3%; Score 28; DB 1; Length 633;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSIGS 6
DB 349 GSIGS 354

RESULT 11
YKD4_YEAST STANDARD: PRT: 758 AA.
AC P36096;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HYPOTHETICAL 87.9 KDA PROTEIN IN PTM-IRX1 INTERGENIC REGION PRECURSOR.
GN YKL034W OR YKL247.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C;
RA Medline=94262309; Pubmed=8203146;
RT "Analysis of an 11.7 kb DNA fragment of chromosome XI reveals a new
RT tRNA gene and four new open reading frames including a leucine zipper
RT protein and a homologue to the yeast mitochondrial regulator ABP2.";
RL Yeast 10:125-130(1994).
RN [2]
RP SEQUENCE OF 1-570 FROM N.A.
RC STRAIN-S288C;
RA Medline=93127731; Pubmed=1481573;
RA Purnelle B., Skala J., van Dyck L., Goffeau A.;
RT "The sequence of a 12 kb fragment on the left arm of yeast chromosome
RT

```

```

RT XI reveals five new open reading frames, including a zinc finger
RT protein and a homolog of the UDP-glucose pyrophosphorylase from
RT potato.";
RN Yeast 8:977-986(1992).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X71622; -; NOT_ANNOTATED_CDS.
DR EMBL: X69584; CAA49298.1; -
DR EMBL: Z28034; CAA81869.1; -
DR PIR: S37855; S37855.
DR SGD: S0001517; YKL034W.
DR InterPro: IPR001841; ZnF_ring.
DR Pfam: PF00097; ZF-C3HC4; 1.
DR SMART: SM00184; RING; 1.
KW Hypothetical protein; Transmembrane; Signal.
FT SIGNAL 1 26
FT CHAIN 27 758
FT TRANSMEM 400 416
FT TRANSMEM 440 456
FT TRANSMEM 461 477
FT TRANSMEM 528 544
FT TRANSMEM 607 623
FT TRANSMEM 638 654
FT TRANSMEM 758 AA; 87879 MW; F29A61E94F96628 CRC64;
SQ SEQUENCE 758 AA; 87879 MW; F29A61E94F96628 CRC64;

Query Match
Best Local Similarity 93.3%; Score 28; DB 1; Length 758;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSIGS 6
DB 522 GSIGS 527

RESULT 12
ATCX_SCHPO STANDARD: PRT: 1402 AA.
ID ATCX_SCHPO
AC Q09891;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE POTENTIAL PHOSPHOLIPID TRANSPORTING ATPASE 1 (EC 3.6.3.13).
GN SPC2AB11.12C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RA Odell C., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THIS MAGNESIUM DEPENDENT ENZYME CATALYZES THE
CC HYDROLYSIS OF ATP COUPLED WITH THE TRANSPORT OF CALCIUM
CC (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
CC (E1-E2 ATPASES). SUBFAMILY IV.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: Z67757; CAAG177.1; -  
DR InterPro: IPR001757; E1-E2-ATPase.  
DR InterPro: IPR001454; Hydrolyase.  
DR Pfam: PF00702; Hydrolyase; 1.  
DR PROSITE: PS00154; APPASE\_E1\_E2; 1.  
KW Hydrolyase; Transmembrane; Phosphorylation; Magnesium; ATP-binding.  
FT TRANSMEM 109 129 POTENTIAL.  
FT TRANSMEM 135 155 POTENTIAL.  
FT TRANSMEM 457 477 POTENTIAL.  
FT TRANSMEM 501 521 POTENTIAL.  
FT TRANSMEM 1066 1086 POTENTIAL.  
FT TRANSMEM 1101 1121 POTENTIAL.  
FT TRANSMEM 1151 1171 POTENTIAL.  
FT TRANSMEM 1193 1213 POTENTIAL.  
FT TRANSMEM 1218 1238 POTENTIAL.  
FT TRANSMEM 1260 1280 POTENTIAL.  
FT MOD RES 569 569 PHOSPHORYLATION (BY SIMILARITY).  
SQ SEQUENCE 1402 AA; 159355 MW; E8495A7598B3020B CRC64;  
  
Query Match 93.3%; Score 28; DB 1; Length 1402;  
Best Local Similarity 83.3%; Pred. No. 4.9e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 GSLGCS 6  
Db 490 GSVGCS 495  
11:1111  
  
RESULT 13  
YUG7\_YEAST STANDARD; PRT; 116 AA.  
ID YUG7\_YEAST  
AC P40365;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE HYPOTHEICAL 12.7 KDA PROTEIN IN SMC3-MBP18 INTERGENIC REGION.  
GN YJL067W OR J1107 OR HRA116.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN=5288C;  
RX MEDLINE=95282514; PubMed=7762302;  
RA Vandenbol M., Durand P., Dion C., Portetelle D., Hliger F.;  
RT "Sequence of a 17.1 kb DNA fragment from chromosome X of  
RT Saccharomyces cerevisiae includes the mitochondrial ribosomal protein  
RT L8.";  
RL Yeast 11:57-60(1995).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: Z34288; CAAB4056.1; -  
DR EMBL: Z49342; CAAB9357.1; -  
DR PIR: S47124; S47124.  
DR SGD: S0003603; YJL067W.  
KW Hypothetical protein; Transmembrane.  
FT TRANSMEM 40 60 POTENTIAL.  
FT TRANSMEM 72 92 POTENTIAL.  
SQ SEQUENCE 116 AA; 12748 MW; 6D5A655ABFA990DF CRC64;

Query Match 90.0%; Score 27; DB 1; Length 116;  
Best Local Similarity 83.3%; Pred. No. 65;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 GSVGCS 6  
Db 27 GSVGCS 32  
11:1111  
  
RESULT 14  
YML3\_THIFE STANDARD; PRT; 149 AA.  
ID YML3\_THIFE  
AC P20089;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 01-AUG-1991 (Rel. 19, Last annotation update)  
DE HYPOTHEICAL 15.6 KDA PROTEIN IN MOBL 3' REGION (ORF 5).  
OS Thibacillus ferrooxidans.  
OC Bacteria; Proteobacteria; gamma subdivision; Acidithiobacillus.  
OX NCBI\_TaxID=920;  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 33020;  
RX MEDLINE=91125140; PubMed=2280689;  
RA Droler M., Zanga P., Lau P.C.K.;  
RT "The mobilization and origin of transfer regions of a Thibacillus  
RT ferrooxidans plasmid: relatedness to plasmids RSF1010 and pSC101.";  
RL MOL. Microbiol. 4:1381-1391(1990).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X52699; CA36929.1; -  
KW Hypothetical protein; Plasmid.  
SQ SEQUENCE 149 AA; 15662 MW; 770E8745E5AC7E1C CRC64;  
  
Query Match 90.0%; Score 27; DB 1; Length 149;  
Best Local Similarity 83.3%; Pred. No. 84;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 GSVGCS 6  
Db 92 GSVGCS 97  
11:1111  
  
RESULT 15  
HIS7\_AZOB STANDARD; PRT; 207 AA.  
ID HIS7\_AZOB  
AC P18787;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE IMIDAOLGLYCEROL-PHOSPHATE DEHYDRATASE (EC 4.2.1.19) (IGPD).  
GN HISB.  
OS Acetivallium brasiliense.  
OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillaceae;  
OC Azospirillum.  
OX NCBI\_TaxID=192;  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN=Sp6;  
RX MEDLINE=89313660; PubMed=2664449;  
RA Paul R., Bazzicalupo M., Damiani G., Bianchi A., Schipani C.,  
RA Sgarbetta V., Polsinelli M.;

```

CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL: Y00459; CA68518.1;
CC  DR PIR: S17776; S17776.
CC  KM Streptomycin biosynthesis.
CC  SQ SEQUENCE 281 AA; 31726 MW; 82456C5119B6387C CRC64;

Query Match
Best Local Similarity 83.3%; Score 27; DB 1; Length 281;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLGS 6
DB 11:111
DB 91 GSVGS 96

RESULT 17
HO3_RAT
ID HO3_RAT STANDARD; PRT; 290 AA.
AC 070453.
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HEME OXYGENASE 3 (EC 1.14.99.3) (HO-3).
HMOX3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RC MEDLINE=97410004; PubMed=9266719;
RA McCoubrey W.K. Jr., Huang T.J., Malnes M.D.;
RT "Isolation and characterization of a cDNA from the rat brain that
RT encodes hemoprotein heme oxygenase-3."
RL Eur. J. Biochem. 247:725-732(1997).
CC -1- FUNCTION: HEME OXYGENASE CLEAVES THE HEME RING AT THE ALPHA
CC METHENE BRIDGE TO FORM BILIVERDIN. BILIVERDIN IS SUBSEQUENTLY
CC CONVERTED TO BILIRUBIN BY BILIVERDIN REDUCTASE. HEME OXYGENASE 3
CC COULD BE IMPLICATED IN SOME HEME-DEPENDENT REGULATORY ROLE IN THE
CC CELL.
CC -1- CATALYTIC ACTIVITY: HEME + 3 AH(2) + O(2) = BILIVERDIN + FE(2+) +
CC CO + 3 A + 3 H(2)O.
CC -1- SUBCELLULAR LOCATION: MICROSOMAL.
CC -1- TISSUE SPECIFICITY: FOUND IN THE SPLEEN, LIVER, THYMUS, PROSTATE,
CC HEART, KIDNEY, BRAIN AND TESTIS.
CC -1- SIMILARITY: BELONGS TO THE HEME OXYGENASE FAMILY.
CC -1- SIMILARITY: CONTAINS 2 HEME REGULATORY MOTIFS (HRM).
CC -----
CC THE SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage is for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF058787; AAC14142.1;
CC DR InterPro: IPR002051; Heme_oxygenase.
CC DR Pfam: PF01126; Heme_oxygenase; 1.
CC DR PROSITE: PS00593; HEME_OXYGENASE; 1.
CC KM Heme; Oxidoreductase; Microsome; Multigene family.
CC FT DOMAIN 238 243 HRM 1 (PROBABLE).
CC FT DOMAIN 255 260 HRM 2 (POTENTIAL).
CC SQ SEQUENCE 290 AA; 32592 MW; 6600235CE8485829 CRC64;

Query Match
Best Local Similarity 83.3%; Score 27; DB 1; Length 290;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 GSLGGS 6  
 DB 249 GTLGGG 254

## RESULT 18

YM82\_MYCTU  
 ID YM82\_MYCTU STANDARD: PRT: 312 AA.  
 AC 030683;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE HYPOTHEICAL TRANSCRIPTIONAL REGULATOR RV2282C.  
 GN RV2282C OR MT2340 OR MTCY339.28.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteriales;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_Taxid=1773;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Tagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Ruston J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolony J.F., Nelson W.C., Umayam L.A., Ermolova M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishal W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains.";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL  
 CC REGULATORS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: 277163; CAB00986.1; -;  
 DR EMBL: AE007077; AAK4624.1; -;  
 DR TIGR: MT2340; -;  
 DR TubercuList: RV2282C; -;  
 DR InterPro: IPR000847; HTH\_LYSR.  
 DR Pfam: PF00126; HTH\_1; 1.  
 DR PRINTS: PR00039; HTH\_LYSR.  
 DR PROSITE: PS00044; HTH\_LYSR\_FAMILY; 1.  
 KW Hypothetical protein; Transcription regulation; DNA-binding;  
 KW Complete proteome.  
 FT DNABIND 25 H-T-H MOTIF (POTENTIAL).  
 FT SEQUENCE 312 AA: 32989 MW: 35299 kDa: 403122 CRC64;

Query Match 90.0%; Score 27; DB 1; Length 312;  
 Best Local Similarity 83.3%; Pred. No. 1.8e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLGGS 6  
 DB 23 GTLGGG 28

## RESULT 19

HO2\_MOUSE  
 ID HO2\_MOUSE STANDARD: PRT: 315 AA.  
 AC 070252; 070626;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE HEME OXYGENASE 2 (EC 1.14.99.3) (HO-2).  
 GN HMOX2.  
 OS Mus musculus (mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_Taxid=10090;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CD-1;  
 RA Mount D.B.;  
 RT "Sequence of mouse heme oxygenase-2.";  
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
 [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99013870; PubMed=9795203;  
 RA Gibbs L., Willis D., Morgan M.J.;  
 RT "The identification and expression of heme oxygenase-2 alternative  
 RT transcripts in the mouse.";  
 RL Gene 221:171-177(1998).  
 CC -1- FUNCTION: HEME OXYGENASE CLEAVES THE HEME RING AT THE ALPHA  
 CC METHENE BRIDGE TO FORM BILIVERDIN. BILIVERDIN IS SUBSEQUENTLY  
 CC CONVERTED TO BILIRUBIN BY BILIVERDIN REDUCTASE. UNDER  
 CC PHYSIOLOGICAL CONDITIONS, THE ACTIVITY OF HEME OXYGENASE IS  
 CC HIGHEST IN THE SPLEEN, WHERE SENESCENT ERYTHROCYTES ARE  
 CC SEQUESTERED AND DESTROYED.  
 CC -1- FUNCTION: HEME OXYGENASE 2 COULD BE IMPLICATED IN THE PRODUCTION  
 CC OF CARBON MONOXIDE IN BRAIN WHERE IT COULD ACT AS A  
 CC NEUROTRANSMITTER.  
 CC -1- CATALYTIC ACTIVITY: HEME + 3 AH(2) + O(2) = BILIVERDIN + FE(2+) +  
 CC CO + 3 A + 3 H(2)O.  
 CC -1- SUBCELLULAR LOCATION: MICROSOMAL.  
 CC -1- SIMILARITY: BELONGS TO THE HEME OXYGENASE FAMILY.  
 CC -1- SIMILARITY: CONTAINS 2 HEME REGULATORY MOTIFS (HRM).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AF029874; AAC17981.1; -;  
 DR EMBL: AF054670; AAC82364.1; -;  
 DR EMBL: AF054669; AAC82363.1; -;  
 DR MGD: MGI:109373; Hmox2.  
 DR InterPro: IPR002051; Heme\_oxygenase.  
 DR Pfam: PF01126; Heme\_oxygenase; 1.  
 DR PRINTS: PR00088; HAEMOXYGNAE.  
 DR PROSITE: PS00593; HEME\_OXYGENASE; 1.  
 KW Heme; Oxidoreductase; Microsome; multigene family.  
 FT BINDING 44 PROXIMAL HEME LIGAND (BY SIMILARITY).  
 FT DOMAIN 263 HRM 1 (POTENTIAL).  
 FT DOMAIN 280 HRM 2 (POTENTIAL).  
 FT CONFLICT 172 RAL -> SSS (IN REF. 2).  
 FT SEQUENCE 315 AA: 35738 MW: 35328 kDa: 403122 CRC64;

Query Match 90.0%; Score 27; DB 1; Length 315;  
 Best Local Similarity 83.3%; Pred. No. 1.8e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSLGGS 6  
1:|||||  
Db 274 GTLGGs 279

RESULT 20  
HO2\_RAT STANDARD; PRT; 315 AA.  
AC P23711:  
DT 01-NOV-1991 (Rel. 20, Created)  
DT 01-NOV-1991 (Rel. 20, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE HEME OXYGENASE 2 (EC 1.14.99.3) (HO-2).  
GN HMOX2.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-Testis;  
RX MEDLINE=90237051; PubMed=2185251;  
RA Rotenberg M.O., Maines M.D.;  
RT "Isolation, characterization, and expression in *Escherichia coli* of a  
RT cDNA encoding rat heme oxygenase-2.";  
RL J. Biol. Chem. 265:7501-7506(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SPRACUE-DAWLEY; TISSUE=Liver;  
RX MEDLINE=94156193; PubMed=8112599;  
RA McCoubrey W.K. Jr., Maines M.D.;  
RT "The structure, organization and differential expression of the gene  
RT encoding rat heme oxygenase-2.";  
RL Gene 139:155-161(1994).  
RN [3]  
RP SEQUENCE OF 142-232 FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE-Testis, and Liver;  
RX MEDLINE=88139412; PubMed=3343248;  
RA Cruse I., Maines M.D.;  
RT "Evidence suggesting that the two forms of heme oxygenase are  
RT products of different genes.";  
RL J. Biol. Chem. 263:3348-3353(1988).  
CC -1- FUNCTION: HEME OXYGENASE CLEAVES THE HEME RING AT THE ALPHA  
CC METHENE BRIDGE TO FORM BILIVERDIN. BILIVERDIN IS SUBSEQUENTLY  
CC CONVERTED TO BILIRUBIN BY BILIVERDIN REDUCTASE. UNDER  
CC PHYSIOLOGICAL CONDITIONS, THE ACTIVITY OF HEME OXYGENASE IS  
CC HIGHEST IN THE SPLEEN, WHERE SENESCENT ERYTHROCYTES ARE  
CC SEQUESTERED AND DESTROYED.  
CC -1- FUNCTION: HEME OXYGENASE 2 COULD BE IMPLICATED IN THE PRODUCTION  
CC OF CARBON MONOXIDE IN BRAIN WHERE IT COULD ACT AS A  
CC NEUOTRANSMITTER.  
CC -1- CATALYTIC ACTIVITY: HEME + 3 AH(2) + O(2) - BILIVERDIN + FE(2+) +  
CC CO + 3 A + 3 H(2)O.  
CC -1- SUBCELLULAR LOCATION: MICROSOAM.  
CC -1- TISSUE SPECIFICITY: WIDELY DISTRIBUTED IN BODY WITH A HIGH  
CC CONCENTRATION IN THE BRAIN.  
CC -1- INDUCTION: HEME OXYGENASE 2 ACTIVITY IS NON-INDUCIBLE.  
CC -1- SIMILARITY: BELONGS TO THE HEME OXYGENASE FAMILY.  
CC -1- SIMILARITY: CONTAINS 2 HEME REGULATORY MOTIFS (HRM).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
CC or send an email to [license@isb-sdb.ch](mailto:license@isb-sdb.ch)).  
CC -----  
CC EMBL: J05405; AAA1340.1; -  
CC EMBL: U05013; AAA19130.1; -  
CC EMBL: M18918; AAA41347.1; -

DR PIR: A29922; A29922.  
DR PIR: A35199; A35199.  
DR InterPro: IPR002051; Heme\_oxygenase.  
DR Pfam: PF01126; Heme\_oxygenase\_1.  
DR PRINTS: PR00088; HAEMOXYGENASE.  
DR PROSITE: PS00593; HEME OXYGENASE: 1.  
KW Heme: Oxidoreductase; Microsome; Multigene family.  
FT BINDING 44 PROXIMAL HEME LIGAND (BY SIMILARITY).  
FT DOMAIN 263 268 HRM 1 (POTENTIAL).  
FT DOMAIN 280 285 HRM 2 (POTENTIAL).  
FT CONFLICT 142 146 ONEPE -> EFFRK (IN REF. 3).  
FT CONFLICT 230 232 MOI -> TEF (IN REF. 3).  
SQ SEQUENCE 315 AA; 35762 MW; 981A9DE01DE1AFCF CRC64;

Query Match 90.0%; Score 27; DB 1; Length 315;  
Best local similarity 83.3%; Pred. No. 1.8e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSLGGS 6  
1:|||||  
Db 274 GTLGGs 279

RESULT 21  
ESTR\_STRSC STANDARD; PRT; 345 AA.  
ID ESTR\_STRSC  
AC P22266;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE ESTERASE PRECURSOR (EC 3.1.1.-).  
GN ESTR.  
OS Streptomyces scabies.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1930;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 40-69.  
RC STRAIN=PL1;  
RX MEDLINE=91072254; PubMed=2254271;  
RA Raymer G., Willard J.M.A., Schottel J.L.;  
RT "Cloning, sequencing, and regulation of expression of an  
RT extracellular esterase gene from the plant pathogen *Streptomyces*  
RT scabies.";  
RL J. Bacteriol. 172:7020-7026(1990).  
RN [2]  
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).  
RX MEDLINE=95292107; PubMed=7773790;  
RA Wei Y., Schottel J.L., Derewenda U., Swenson L., Parker S.,  
RA Derewenda Z.S.;  
RT "A novel variant of the catalytic triad in the Streptomyces scabies  
RT esterase.";  
RL Nat. Struct. Biol. 2:218-223(1995).  
CC -1- INDUCTION: BY ZINC.  
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-5 IS THE INITIATOR.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
CC or send an email to [license@isb-sdb.ch](mailto:license@isb-sdb.ch)).  
CC -----  
CC EMBL: M57297; AAA26743.1; -  
CC EMBL: M57297; AAA26744.1; ALT\_INIT.  
CC PIR: A37845; A37845.  
CC PDB: IESC; 15-OCT-95.  
CC PDB: IESD; 15-OCT-95.  
CC PDB: IESB; 15-OCT-95.  
DR Hydroxylase; Serine esterase; Signal; 3D-structure.  
FT SIGNAL 1 39

FT CHAIN 40 345 ESTERASE.  
 FT DISULFID 73 103  
 FT DISULFID 156 180  
 FT DISULFID 236 294  
 SQ SEQUENCE 345 AA; 36670 MW; 237DC28E185B00E CRC64;

Query Match 90.0%; Score 27; DB 1; Length 345;  
 Best Local Similarity 83.3%; Pred. No. 1.9e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLGGS 6  
 DB 140 GSLGCG 145

RESULT 22  
 NAGA\_VIBCH STANDARD; PRT; 378 AA.  
 AC 032445; Q9KTA9;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE N-ACETYLGLUCOSAMINE-6-PHOSPHATE DEACETYLASE (EC 3.5.1.25) (GLCNAC-6-P DEACETYLASE).  
 GN NAGA OR VC0994.  
 OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=666;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-NON-O1 / 1148A;  
 RX MEDLINE=97446530; PubMed=9301118;  
 RA Yamano N., Oura N., Wang J., Fujishima S.;  
 RT "Cloning and sequencing of the genes for N-acetylglucosamine use that  
 RT construct divergent operons (nagE-nagAC) from Vibrio cholerae  
 RT non-O1.";  
 RL Biochem. Biotechnol. Blochem. 61:1349-1353(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-EL TOR N16961 / SEROTYPE O1;  
 RX MEDLINE=20406833; PubMed=10952301;  
 RA Heidelberg J.F., Eilen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,  
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
 RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragol J., Sellers P.,  
 RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,  
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 RA Fraser C.M.;  
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
 RT cholerae.";  
 RL Nature 406:477-483(2000).  
 CC -1- CATALYTIC ACTIVITY: N-ACETYL-D-GLUCOSAMINE 6-PHOSPHATE + H(2)O  
 CC -1- PATHWAY: N-ACETYL GLUCOSAMINE UTILIZATION PATHWAY.  
 CC -1- SIMILARITY: BELONGS TO THE NAGA FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D87820; BAA22834.1; -;  
 CC EMBL: AE004181; AAF94155.1; -;  
 CC TIGR: VC0994; -;  
 CC InterPro: IPR003764; NAGA.  
 DR Pfam: PF02612; NAGA.1.  
 KW Hydrolase; Carbohydrate metabolism; Complete proteome.  
 FT CONFLICT 79 79 T -> I (IN REF. 1).  
 SQ SEQUENCE 378 AA; 40956 MW; 34906344A3F92A0F CRC64;

Query Match 90.0%; Score 27; DB 1; Length 378;  
 Best Local Similarity 83.3%; Pred. No. 2.1e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLGGS 6  
 DB 301 GTLGGS 306

RESULT 23  
 PGLR\_TOBAC STANDARD; PRT; 396 AA.  
 ID PGLR\_TOBAC  
 AC 005967;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE POLYGALACTURONASE PRECURSOR (EC 3.2.1.15) (PG) (PECTINASE).  
 GN PGL.  
 OS Nicotiana tabacum (Common tobacco).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; eusterids I; Solanales; Solanaceae; Nicotiana.  
 OX NCBI\_TaxID=4097;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. HAVANA; TISSUE=Pollen;  
 RX MEDLINE=9428951; PubMed=8018876;  
 RA Tebbutt S.J., Lonsdale D.M.;  
 RT "Characterization of a tobacco gene encoding a pollen-specific  
 RT polygalacturonase.";  
 RL Plant Mol. Biol. 25:283-297(1994).  
 CC -1- FUNCTION: MAY FUNCTION IN DEPOLYMERIZING PECTIN DURING POLLEN  
 CC DEVELOPMENT, GERMINATION, AND TUBE GROWTH.  
 CC -1- CATALYTIC ACTIVITY: RANDOM HYDROLYSIS OF 1,4-ALPHA-D-  
 CC GALACTOSIDURONIC LINKAGES IN PECTATE AND OTHER GALACTURONANS.  
 CC -1- SUBCELLULAR LOCATION: SECRETED.  
 CC -1- TISSUE SPECIFICITY: POLLEN.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES  
 CC (POLYGALACTURONASES).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X71020; CAA50338.1; -;  
 CC EMBL: X71017; CAA50335.1; ALT\_INT.  
 CC EMBL: X71018; CAA50336.1; ALT\_INT.  
 CC EMBL: X71016; CAA50334.1; ALT\_INT.  
 CC EMBL: X71019; CAA50337.1; ALT\_INT.  
 CC PIR: S32007; S32007.  
 DR InterPro: IPR000743; Polygalacturonase.  
 DR Pfam: PF00295; Glyco\_hydro\_28; 1.  
 DR PROSITE: PS00502; POLYGALACTURONASE; 1.  
 KW Hydrolase; Glycosidase; Cell wall; Signal; Glycoprotein.  
 FT SIGNAL 1 22  
 FT CHAIN 23 396  
 FT ACT\_SITE 236 236 PROBABLE.  
 FT CARBOHYD 143 143 N-LINKED (GLCNAC...)  
 FT CARBOHYD 151 151 N-LINKED (GLCNAC...)  
 FT CARBOHYD 174 174 N-LINKED (GLCNAC...)  
 FT CARBOHYD 181 181 N-LINKED (GLCNAC...)  
 FT CARBOHYD 203 203 N-LINKED (GLCNAC...)  
 FT CARBOHYD 208 208 N-LINKED (GLCNAC...)  
 FT CARBOHYD 259 259 N-LINKED (GLCNAC...)  
 FT CARBOHYD 294 294 N-LINKED (GLCNAC...)  
 FT VARIANT 154 154 I -> V.  
 FT VARIANT 258 258 R -> S.



Db 331 GSVGS 336

RESULT 25

Y4XM\_RHISN STANDARD: PRT: 404 AA.

AC P55705;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE HYPOTHETICAL TRANSPORT PROTEIN Y4XM.

GN Y4XM.

OS Rhizobium sp. (strain NGR234).

OC Plasmid sym pNGR234.

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Rhizobiaceae; Rhizobium.

OX NCBI\_TaxID=394;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE-97305956; PubMed-9163424;

RA Freiberg C.A., Fellay R., Balroch A., Broughton W.J., Rosenthal A., Perret X.;

RT "Molecular basis of symbiosis between Rhizobium and legumes.";

RL Nature 387:394-401(1997).

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE (POTENTIAL).

CC -1- SIMILARITY: TO E.COLI YCEE. ALSO TO TETRACYCLINE TRANSPORTERS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; AE000106; AAB91936.1; -

DR InterPro: IPR003662; sub-transporter.

DR Pfam: PF00083; sugar\_tr: 1.

KW Hypothetical protein; Transport; Transmembrane; Inner membrane;

KW Plasmid.

FT TRANSMEM 11 31 POTENTIAL.

FT TRANSMEM 49 69 POTENTIAL.

FT TRANSMEM 89 109 POTENTIAL.

FT TRANSMEM 151 171 POTENTIAL.

FT TRANSMEM 178 198 POTENTIAL.

FT TRANSMEM 211 231 POTENTIAL.

FT TRANSMEM 253 273 POTENTIAL.

FT TRANSMEM 284 304 POTENTIAL.

FT TRANSMEM 311 331 POTENTIAL.

FT TRANSMEM 346 366 POTENTIAL.

FT TRANSMEM 377 397 POTENTIAL.

SO SEQUENCE 404 AA; 42013 MW; C243456E33730F31 CRC64;

QY 1 GSLSGS 6

Db 152 GSLGCA 157

Query Match 90.0%; Score 27; DB 1; Length 404;

Best Local Similarity 83.3%; Pred. No. 2.3e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Y 1 GSLSGS 6

YCDG\_ECOLI

ID YCDG\_ECOLI STANDARD: PRT: 442 AA.

AC P75892; O9R3W5;

DT 01-NOV-1997 (Rel. 35, Created)

DT 20-AUG-2001 (Rel. 40, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE PUTATIVE PURINE PERMEASE YCDG.

GN YCDG OR B1006.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

OX NCBI\_TaxID=562;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-K12 / MG1655;

RX MEDLINE-97426617; PubMed-9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Colado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.O., Mau B., Shao Y.;

RT "The complete genome sequence of Escherichia coli K-12.";

RL Science 277:1453-1474(1997).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-K12;

RX MEDLINE-97061202; PubMed-8905232;

RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A., Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K., Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N., Sempel G., Seki Y., Tagami H., Takemoto K., Mada C., Yamamoto Y., Yano M., Horiiuchi T.;

RT "A 7.18-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 12.7-28.0 min region on the linkage map.";

RL DNA Res. 3:137-155(1996).

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).

CC -1- SIMILARITY: BELONGS TO THE XANTHINE/URACIL PERMEASES FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; AE000202; AAC74091.1; ALT\_INIT.

DR EMBL; D90737; BAA35773.1; -

DR EMBL; D90738; BAA35783.1; -

DR Ecogene; EG18853; ycdG.

DR InterPro: IPR000444; Xan\_ur\_permease.

DR Pfam: PF00860; xan\_ur\_permease; 1.

DR PROSITE: PS01116; XANTH URACIL PERMEASE; 1.

KW Hypothetical protein; Transmembrane; Transport; Complete proteome.

FT TRANSMEM 58 78 POTENTIAL.

FT TRANSMEM 80 100 POTENTIAL.

FT TRANSMEM 111 131 POTENTIAL.

FT TRANSMEM 141 161 POTENTIAL.

FT TRANSMEM 170 190 POTENTIAL.

FT TRANSMEM 197 217 POTENTIAL.

FT TRANSMEM 241 261 POTENTIAL.

FT TRANSMEM 285 305 POTENTIAL.

FT TRANSMEM 319 339 POTENTIAL.

FT TRANSMEM 348 368 POTENTIAL.

FT TRANSMEM 386 406 POTENTIAL.

FT TRANSMEM 407 427 POTENTIAL.

SO SEQUENCE 442 AA; 45557 MW; 35CC612A1E86D67 CRC64;

QY 1 GSLSGS 6

Db 297 GSVGS 302

Query Match 90.0%; Score 27; DB 1; Length 442;

Best Local Similarity 83.3%; Pred. No. 2.5e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Y 1 GSLSGS 6

MSRE\_MOUSE

ID MSRE\_MOUSE STANDARD: PRT: 458 AA.







CC -1- SIMILARITY: CONTAINS 3 SRCR DOMAINS.  
CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD5 entry;  
CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd5.htm".  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X04391: CAA27979.1: -  
DR EMBL: X89405: CAA61584.2: -  
DR EMBL: AJ237927: CAA61584.2: JOINED.  
DR EMBL: AJ237928: CAA61584.2: JOINED.  
DR EMBL: AJ237929: CAA61584.2: JOINED.  
DR EMBL: AJ237930: CAA61584.2: JOINED.  
DR EMBL: AJ237931: CAA61584.2: JOINED.  
DR EMBL: AJ237932: CAA61584.2: JOINED.  
DR PIR: A26396: A26396.  
DR GlycoSuiteDB: P06127: -  
DR MIM: 153340: -  
DR InterPro: IPR001190: SRCR.  
DR Pfam: PF00530: SRCR. 3.  
DR SMART: SM00202: SR. 2.  
DR PROSITE: PS00420: SRCR\_1; FALSE\_NEG.  
DR PROSITE: PS50287: SRCR\_2; 3.  
KW Signal: Transmembrane; Glycoprotein; T-cell; Repeat.  
FT SIGNAL 1 24  
FT CHAIN 25 495 T-CELL SURFACE GLYCOPROTEIN CD5.  
FT DOMAIN 25 372 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 373 402 POTENTIAL.  
FT DOMAIN 403 495 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 35 133 SRCR 1.  
FT DOMAIN 159 268 SRCR 2.  
FT DOMAIN 276 368 SRCR 3.  
FT DISULFID 44 107 POTENTIAL.  
FT CARBOHD 116 116 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHD 241 241 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SO SEQUENCE 495 AA: 54625 MW: 8740DD7623DE1D3 CRC64;  
  
Query Match 90.0%; Score 27; DB 1; Length 495;  
Best Local Similarity 83.3%; Pred. No. 2.8e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 GSLGGS 6  
DB 179 GSLGGS 184  
  
RESULT 31  
HDAC\_ARATH STANDARD: PRT; 501 AA.  
AC 022446;  
ID HDAC\_ARATH STANDARD: PRT; 501 AA.  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE HISTONE DEACETYLASE (HD).  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eustosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_Taxid=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA Tomihama T., Shoji K., Hanyu H., Okano T.;  
RE Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: RESPONSIBLE FOR THE DEACETYLATION OF LYSINE RESIDUES ON  
CC THE N-TERMINAL PART OF THE CORE HISTONES (H2A, H2B, H3 AND H4).  
CC HISTONE DEACETYLATION PLAYS AN IMPORTANT ROLE IN TRANSCRIPTIONAL

CC REGULATION, CELL CYCLE PROGRESSION AND DEVELOPMENTAL EVENTS (BY  
CC SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE HISTONE DEACETYLASE / ACUC / ALPHA  
CC FAMILY. HD SUBFAMILY 1.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AF014824: AAB66486.1: -  
DR InterPro: IPR00286: Hist\_deacetylase.  
DR Pfam: PF00850: Hist\_deacetyl; 1.  
DR PRINTS: PR01270: HDASUPER.  
DR PRINTS: PR01271: HISDACETLASE.  
KW Hydrolase; Nuclear protein.  
SO SEQUENCE 501 AA: 56037 MW: C50AF5624958D6C2 CRC64;  
  
Query Match 90.0%; Score 27; DB 1; Length 501;  
Best Local Similarity 83.3%; Pred. No. 2.8e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 GSVGGS 6  
DB 124 GSVGGS 129  
  
RESULT 32  
USP\_DROME STANDARD: PRT; 508 AA.  
ID USP\_DROME STANDARD: PRT; 508 AA.  
AC P20153: Q9W535;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE ULTRASPIRACLE PROTEIN (XR2C) (CHORION FACTOR 1).  
GN USP OR NR2B4 OR CFI OR EG:22E5.1 OR CC4380.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_Taxid=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Larva;  
RX MEDLINE-90384571: PubMed-2169594;  
RA Oro A.E., McKewen M., Evans R.M.;  
RT "Relationship between the product of the Drosophila ultraspiracle  
RT locus and the vertebrate retinoid X receptor.";  
RL Nature 347:298-301(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CANTON-S;  
RX MEDLINE-90332404: PubMed-2165589;  
RA Heinrich V.C., Silter T.J., Lubahn D.B., MacIntyre A., Gilbert L.I.;  
RT "A steroid/thyroid hormone receptor superfamily member in Drosophila  
RT melanogaster that shares extensive sequence similarity with a  
RT mammalian homolog.";  
RL Nucleic Acids Res. 18:4143-4148(1990).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OREGON-R;  
RX MEDLINE-20196011: PubMed-10731137;  
RA Benos P.V., Gatti M.K., Ashburner M., Murphy L., Harris D.,  
RA Barrell B.G., Ferraz C., Vidal S., Brun C., Demailles J., Cadieu E.,  
RA Dreano S., Gloux S., Lelaure V., Mottler S., Galibert F., Botkova D.,  
RA Minina B., Katalos F.C., Louis C., Siden-Kiamos I., Bolshakov S.,  
RA Papagiamakis G., Spanos L., Cox S., Madueno E., de Pablo B.,  
RA Modolell J., Peter A., Schottler P., Werner M., Mourikotli F.,

RA Belnert N., Dove G., Schaefer U., Jaekle H., Bucheton A.,  
 RA Callister D.M., Campbell L.A., Darlatisou A., Henderson N.S.,  
 RA McMillan P.J., Salles C., Tait E.A., Valentini P., Saunders R.D.C.,  
 RA Glover D.M.;  
 RT "From sequence to chromosome: the tip of the X chromosome of D.  
 RT melanogaster.";  
 RL Science 287:2220-2222(2000).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY:  
 RX MEDLINE-20196006; PubMed-10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers J.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Baldwin D.,  
 RA Beeson K.Y., Benson P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
 RA Burla K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Paulis B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fowler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idagwam C.,  
 RA Jaitell M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclib J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.C.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E.C., Spindler A.C., Stapleton M., Strong R., Sun E.,  
 RA Sytkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveril J.S., Zhan M., Zhang Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 RN [5]  
 RP SEQUENCE OF 15-294 FROM N.A.  
 RC TISSUE-Ovary;  
 RX MEDLINE-91007257; PubMed-2120114;  
 RA Shea M.J., King D.L., Conboy M.J., Mariani B.D., Kafatos F.C.;  
 RT "Proteins that bind to *Drosophila* chorion cis-regulatory elements: a  
 RT new C2H2 zinc finger protein and a C2C2 steroid receptor-like  
 RT component.";  
 RL Genes Dev. 4:1128-1140(1990).  
 RN [6]  
 RP SUBUNITS  
 RX MEDLINE-94067348; PubMed-8247157;  
 RA Yao T.-P., Froman B.M., Jiang Z., Cherbas L., Chen J.-D.,  
 RA McKown M.M., Cherbas P., Evans R.M.;  
 RT "Functional ecdysone receptor is the product of *Ecr* and *Ultraspindle*  
 RT genes.";  
 RL Nature 366:476-479(1993).  
 CC -1- FUNCTION: RECEPTOR FOR ECDYSONE. MAY BE AN IMPORTANT MODULATOR OF  
 CC INSECT METAMORPHOSIS. PLAYS AN IMPORTANT PART IN EMBRYONIC AND  
 CC POST-EMBRYONIC DEVELOPMENT. BINDS TO ECDYSONE RESPONSE ELEMENTS  
 CC (ECRES) SUCH AS IN THE PROMOTER REGION OF S15 CHORION GENE.  
 CC -1- SUBUNIT: HETERODIMER OF USP AND ECR. ONLY THE HETERODIMER IS  
 CC CAPABLE OF HIGH-AFFINITY BINDING TO ECDYSONE.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.

CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,  
 CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.  
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.  
 CC NR2 SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X53417; CAA37496.1; -;  
 DR EMBL: X52591; CAA36827.1; -;  
 DR EMBL: AL031765; CAA21122.1; -;  
 DR EMBL: AE003422; AAF45707.1; -;  
 DR EMBL: X53379; CAA37459.1; -;  
 DR PIR: A35872; A35872.  
 DR PIR: S11513; S11513.  
 DR PIR: S13119; S13119.  
 DR HSSP: P19793; 2NLL.  
 DR TRNSPAC: T00117; -;  
 DR FlyBase: FBgn0003964; usp.  
 DR InterPro: IPR000536; Hormone\_rec\_1lg.  
 DR InterPro: IPR001723; Steroidhormone\_rcptor.  
 DR InterPro: IPR001626; zf-C4.  
 DR Pfam: PF00104; hormone\_rec\_1.  
 DR Pfam: PF00105; zf-C4; 1.  
 DR PRINTS: PRO0047; STROIDFINGER.  
 DR PRINTS: PRO0398; STROIDHORMONER.  
 DR SMART: SM00430; HOL1; 1.  
 DR SMART: SM00399; znf-C4; 1.  
 DR PROSITE: PS00031; NUCLEAR\_RECEPTOR; 1.  
 DR Receptor: Transcription regulation; DNA-binding; Nuclear protein;  
 DR Zinc-finger.  
 FT DOMAIN 1 103 MODULATING (BY SIMILARITY).  
 FT ZN\_BIND 104 169 NUCLEAR RECEPTOR-TYPE.  
 FT ZN\_FING 104 124 C4-TYPE.  
 FT ZN\_FING 140 164 C4-TYPE.  
 FT DOMAIN 170 223 LIGAND.  
 FT DOMAIN 224 508 LIGAND-BINDING (BY SIMILARITY).  
 FT CONFLICT 213 213 G -> E (IN REF. 2).  
 FT CONFLICT 339 339 MISSING (IN REF. 2).  
 SQ SEQUENCE 508 AA; 55244 MW; 58BA7A9D9DEB80 CRC64;  
 Query Match 90.0%; Score 27; DB 1; Length 508;  
 Best Local Similarity 83.3%; Pred. No. 2.8e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GSIGS 6  
 DB 200 GSVGS 205  
 RESULT 33  
 P69\_MYCHR STANDARD; PRT; 560 AA.  
 AC P15362;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE TRANSPORT SYSTEM PERMEASE PROTEIN p69.  
 OS Mycoplasma hyorhinis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;  
 OC Mycoplasmataceae; Mycoplasma.  
 OX NCBI\_TaxID=2100;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-89091146; PubMed-3208756;  
 RA Dudler R., Schmidhauser C., Parish R.W., Wettenhall R.E.H.,  
 RA Schmidt T.;

RT "A mycoplasma high-affinity transport system and the in vitro  
RL invasiveness of mouse sarcoma cells.";  
CC EMBL J. 7:3963-3970(1988).  
CC -1- FUNCTION: PROBABLY PART OF A HIGH-AFFINITY TRANSPORT SYSTEM.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
CC -1- SIMILARITY: WITH INTEGRAL MEMBRANE COMPONENTS OF OTHER BINDING-  
CC PROTEIN-DEPENDENT TRANSPORT SYSTEMS. CONTAINS TWO DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: X14140; CAA32358.1; -;  
DR EMBL: M37339; AAA25429.1; -;  
DR PIR: S01830; S01830.  
DR InterPro: IPR000515; BPD\_TRANSP.  
DR PROSITE: PS00402; BPD\_TRANSP\_INN\_MEMBER; FALSE\_NEG.  
DR TransPort: Transmembrane; Repeat.  
KW TRANSMEM 22 42 POTENTIAL.  
FT TRANSMEM 83 103 POTENTIAL.  
FT TRANSMEM 120 140 POTENTIAL.  
FT TRANSMEM 150 170 POTENTIAL.  
FT TRANSMEM 214 234 POTENTIAL.  
FT TRANSMEM 247 267 POTENTIAL.  
FT TRANSMEM 295 315 POTENTIAL.  
FT TRANSMEM 349 369 POTENTIAL.  
FT TRANSMEM 403 423 POTENTIAL.  
FT TRANSMEM 464 484 POTENTIAL.  
FT TRANSMEM 490 - 510 POTENTIAL.  
FT TRANSMEM 527 547 POTENTIAL.  
FT REPEAT 13 280 APPROXIMATE.  
FT REPEAT 281 547 APPROXIMATE.  
SQ SEQUENCE 580 AA; 69233 MW; 7FE32C044EDD4749 CRC64;  
  
Query Match 90.0%; Score 27; DB 1; Length 580;  
Best Local Similarity 83.3%; Pred. No. 3.3e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GSIGS 6  
DB 495 GSVGS 500  
  
RESULT 34  
TBFA\_EPRST STANDARD; PRT; 643 AA.  
ID TBFA\_EPRST  
AC Q90501;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE THREAD BIOPOLYMER FILAMENT ALPHA SUBUNIT.  
OS Eptaretus stoueti (Pacific hagfish).  
OC Eukaryota; Metazoa; Chordata; Craniata; Hyperotreti; Myxiniiformes;  
OC Myxiniidae; Eptaretinae; Eptaretus.  
OX NCBI\_TaxID=7765;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE=Slime gland;  
RX MEDLINE=9521315; PubMed=7535307;  
RA Koch E.A., Spitzer R.H., Pithawalla R.B., Parry D.A.D.;  
RT "An unusual intermediate filament subunit from the cytoskeletal  
RT biopolymer released extracellularly into seawater by the primitive  
RT hagfish (Eptaretus stoueti).";  
RL J. Cell Sci. 107:3133-3144(1994).  
CC -1- FUNCTION: RELEASED EXTRACELLULARLY INTO SEAWATER AND PROVIDES  
CC PHYSICAL AND BIOLOGICAL DEFENSE AGAINST INVASIVE ORGANISMS BY  
CC MODULATION OF THE VISCOCOLASTIC PROPERTIES OF MUCUS.  
CC -1- SUBUNIT: COILED-COIL HETERODIMER OF AN ALPHA AND A GAMMA SUBUNIT.

CC ASSEMBLE INTO 10 NM FILAMENTS. FORMS A MASSIVE, CONICAL,  
CC INTERMEDIATE FILAMENT BIOPOLYMER OF APPROXIMATELY 60 CM.  
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.  
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: U11865; AAC59661.1; -;  
DR InterPro: IPR001664; IF.  
DR Pfam: PF00038; filament; 2.  
DR PROSITE: PS00266; IF; 1.  
KW Intermediate filament; Coiled coil.  
FT DOMAIN 1 191 HEAD.  
FT DOMAIN 193 510 ROD.  
FT DOMAIN 511 643 TAIL.  
FT DOMAIN 193 227 COIL 1A.  
FT DOMAIN 228 240 LINKER 1.  
FT DOMAIN 241 341 COIL 1B.  
FT DOMAIN 342 362 LINKER 12.  
FT DOMAIN 363 381 COIL 2A.  
FT DOMAIN 382 389 LINKER 2.  
FT DOMAIN 390 510 COIL 2B.  
FT DOMAIN 563 590 POLY-GLY.  
SQ SEQUENCE 643 AA; 66694 MW; FB146930BAFD819C CRC64;  
  
Query Match 90.0%; Score 27; DB 1; Length 643;  
Best Local Similarity 83.3%; Pred. No. 3.6e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GSIGS 6  
DB 68 GSVGS 73  
  
RESULT 35  
FXO3\_HUMAN STANDARD; PRT; 673 AA.  
ID FXO3\_HUMAN  
AC O43524; O15171;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE FORKHEAD PROTEIN O3A (FORKHEAD IN RHABDOMYOSARCOMA-LIKE 1) (AF6Q21  
DE PROTEIN).  
OS FOXO3A OR FKHL1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98140118; PubMed=9479491;  
RA Anderson M.J., Viars G.S., Czeckay S., Caveness W.K., Arden K.C.;  
RT "Cloning and characterization of three human forkhead genes that  
RT comprise an FKHR-like gene subfamily.";  
RL Genomics 47:187-199(1998).  
RN [2]  
RP SEQUENCE OF 1-383 FROM N.A.  
RX MEDLINE=98008138; PubMed=9345057;  
RA Hillion J., Le Contat M., Jonveaux P., Berger R., Bernard O.A.;  
RT "Afeq21, a novel partner of the MLL gene in t(6;11)(q21;q23), defines  
RT a forkhead transcriptional factor subfamily.";  
RL Blood 90:3714-3719(1997).  
CC -1- FUNCTION: PROBABLE TRANSCRIPTION FACTOR THAT MAY TRIGGERS  
CC APOPTOSIS BY INDUCING THE EXPRESSION OF GENES THAT ARE CRITICAL  
CC FOR CELL DEATH.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.

```

CC -1- TISSUE SPECIFICITY: UBIQUITOUS.
CC -1- PTM: PHOSPHORYLATED BY AKT1.
CC -1- DISEASE: INVOLVED IN A T(6;11)(Q21;Q23) CHROMOSOMAL TRANSLOCATION
CC -1- IN SECONDARY ACUTE LEUKEMIA THAT INVOLVES FOXO3A AND MLL.
CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF032886; AAC39592.1; -
DR EMBL: AJ001589; CAA04860.1; -
DR EMBL: AJ001590; CAA04861.1; -
DR HSP: Q63245; 2HFH.
DR MIM: 602681; -
DR InterPro: IPR001766; Fork_head.
DR Pfam: PF00250; Fork_head.1.
DR PRINTS: PR00053; FORKHEAD.
DR SMART: SM00339; FH.1.
DR PROSITE: PS00657; FORK_HEAD_1; FALSE_NEG.
DR PROSITE: PS00658; FORK_HEAD_2; 1.
DR PROSITE: PS50039; FORK_HEAD_3; 1.
DR K1M Transcription regulation: DNA-binding; Nuclear protein; Apoptosis;
KM Chromosomal translocation; Proto-oncogene; Phosphorylation.
FT CONFLICT 156 163
FT CONFLICT 238 246
FT CONFLICT 253 253
FT CONFLICT 271 271
FT CONFLICT 292 330
FT FT
FT FT
FT CONFLICT 345 361
FT CONFLICT 367 367
FT CONFLICT 371 371
FT CONFLICT 382 383
SO SEQUENCE 673 AA: 71276 MW: 554853065A9982 CRC64;

Query Match
Best Local Similarity 83.3%; Score 27; DB 1; Length 673;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSIGGS 6
DB 542 GALS GS 547

RESULT 36
FATA_VIBAN
AC P11461; P19830; STANDARD; PRT; 726 AA.
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE FERRIC ANGIUBACTIN RECEPTOR PRECURSOR (OM2).
GN FATA.
OS Vibrio anguillarum (Listonella anguillarum).
OC Plasmid pJM1.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Listonella.
OX NCBI_Taxid=55601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=775;
RX MEDLINE=88139336; PubMed=2830268;
RA Actis L.A., Tolmashy M.E., Farrell D.H., Croa J.H.;
RT "Genetic and molecular characterization of essential components of
the Vibrio anguillarum plasmid-mediated iron-transport system.";
RL J. Biol. Chem. 263:2853-2860(1988).

```

```

RN [12]
RP SEQUENCE OF 637-726 FROM N.A.
RC STRAIN=775;
RX MEDLINE=90185247; PubMed=2311935;
RA Farrell D.H., Mikesell P., Actis L.A., Croa J.H.;
RT "A regulatory gene, angr, of the iron uptake system of Vibrio
anguillarum: similarity with phage p22 cro and regulation by Iron.";
RL Gene 86:45-51(1990).
CC -1- FUNCTION: ANGIUBACTIN RECEPTOR, ESSENTIAL COMPONENT OF THE IRON
UPAKE SYSTEM OF V. ANGIUBACTIN.
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J03529; AAA91581.1; -
DR EMBL: M34504; AAA79859.1; -
DR PIR: B29928; B29928.
DR PIR: P00051; P00051.
DR InterPro: IPR000531; Tonb_boxc.
DR Pfam: PF000593; Tonb_boxc.1.
DR PROSITE: PS00430; TONB_DEPENDENT_REC_1; 1.
DR PROSITE: PS01156; TONB_DEPENDENT_REC_2; 1.
KM Signal; Receptor; Transmembrane; Outer membrane; Iron transport;
KW Transport; Tonb box; Plasmid.
FT SIGNAL 1 35
FT CHAIN 36 726
FT SITE 49 56
FT SITE 709 726
FT SITE TONB C-TERMINAL BOX.
SO SEQUENCE 726 AA: 78894 MW: 88867A219395B154 CRC64;

Query Match
Best Local Similarity 83.3%; Score 27; DB 1; Length 726;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSIGGS 6
DB 178 GSVGS 183

RESULT 37
PIGR_RABIT
ID PIGR_RABIT STANDARD; PRT; 773 AA.
AC P01832;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE POLYMERIC-IMMUNOGLOBULIN RECEPTOR PRECURSOR (POLY-IG RECEPTOR) (PIGR)
DE [CONTAINS: SECRETORY COMPONENT].
GN PIGR.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_Taxid=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=84142246; PubMed=6322002;
RA Mostov K.E., Friedlander M., Blobel G.;
RT "The receptor for transepithelial transport of IgA and Igm contains
multiple immunoglobulin-like domains.";
RL Nature 308:37-43(1984).
RN [2]
RP SEQUENCE OF 87-114 AND 410-428
RX MEDLINE=88228032; PubMed=313139;
RA Fritliger S., Hughes G.J., Hanly W.C., Jaton J.-C.;
RT "Rabbit secretory components of different allotypes vary in their

```

carbohydrate content and their sites of N-linked glycosylation.";  
 RL J. Biol. Chem. 263:8120-8125(1988).  
 CC -I- FUNCTION: THIS RECEPTOR BINDS POLYMERIC IGA AND IGM AT THE  
 CC BASOLATERAL SURFACE OF EPITHELIAL CELLS. THE COMPLEX IS THEN  
 CC TRANSPORTED ACROSS THE CELL TO BE SECRETED AT THE APICAL SURFACE.  
 CC DURING THIS PROCESS A CLEAVAGE OCCURS THAT SEPARATE THE  
 CC EXTRACELLULAR (KNOWN AS THE SECRETORY COMPONENT) FROM THE  
 CC TRANSMEMBRANE SEGMENT.  
 CC -I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ALSO SECRETED.  
 CC -I- POLYMORPHISM: THE SEQUENCE SHOWN IS THAT OF ALLOTYPIC T62.  
 CC -I- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE V-TYPE DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@sdb-sib.ch](mailto:license@sdb-sib.ch)).  
 CC -----  
 DR EMBL: X00412; CAA25118.1; -  
 DR PIR: A02111; ORRG.  
 DR PIR: A28077; A28077.  
 DR InterPro: IPR003599; 19.  
 DR InterPro: IPR003006; 19\_MHC.  
 DR InterPro: IPR003600; 19\_Like.  
 DR Pfam: PF00047; 19; 5.  
 DR SMART: SM00410; IG\_Like; 1.  
 DR Immunoglobulin domain; Repeat; Transmembrane; Glycoprotein; Signal;  
 KM Polymorphism.  
 FT SIGNAL 1 18  
 FT CHAIN 19 773  
 FT CHAIN 19 615  
 FT DOMAIN 19 647  
 FT TRANSMEM 648 670  
 FT DOMAIN 671 773  
 FT DOMAIN 30 136  
 FT DOMAIN 137 243  
 FT DOMAIN 244 350  
 FT DOMAIN 351 456  
 FT DOMAIN 457 558  
 FT DISULFID 46 115  
 FT DISULFID 155 225  
 FT DISULFID 260 324  
 FT DISULFID 369 438  
 FT DISULFID 478 538  
 FT CARBOHYD 88 108  
 FT CARBOHYD 108 108  
 FT CARBOHYD 418 418  
 FT VARIANT 88 88  
 FT VARIANT 94 94  
 FT VARIANT 101 108  
 FT VARIANT 110 110  
 FT SEQUENCE 773 AA; 83886 MW; DF2C44D2F1193C65 CRC64;  
 Query Match Score 27; DB 1; Length 773;  
 Best Local Similarity 83.3%; Pred. No. 4.3e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

20-AUG-2001 (rel. 40, last annotation update)  
 DE PEROXISOMAL HYDRATASE-DEHYDROGENASE-EPIMERASE (HDE) (MULTIFUNCTIONAL  
 DE BETA-OXIDATION PROTEIN) (MP) [INCLUDES: 2-ENOYL-COA HYDRATASE  
 DE (EC 4.2.1.-); D-3-HYDROXYACYL COA DEHYDROGENASE (EC 1.1.1.-)].  
 GN FOX2 OR YKR009C OR YK108.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_Taxid=4932;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C;  
 RX MEDLINE=93070612; PubMed=1441752;  
 RA Duesterhoeft A., Philippson P.;  
 RT "DNA sequencing and analysis of a 24.7 kb segment encompassing  
 RT centromere CEN11 of Saccharomyces cerevisiae reveals nine previously  
 RT unknown open reading frames.";  
 RL J. Biol. Chem. 267:6646-6653(1992).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92202210; PubMed=1551874;  
 RA Hiltunen J.K., Wenzel B., Beyer A., Erdmann R., Fossa A., Kunau W.H.;  
 RT "Peroxisomal multifunctional beta-oxidation protein of Saccharomyces  
 RT cerevisiae. Molecular analysis of the fox2 gene and gene product.";  
 RL J. Biol. Chem. 267:6646-6653(1992).  
 CC -I- FUNCTION: SECOND TRIFUNCTIONAL ENZYME ACTING ON THE BETA-OXIDATION  
 CC PATHWAY FOR FATTY ACIDS, POSSESSING HYDRATASE-DEHYDROGENASE-  
 CC EPIMERASE ACTIVITIES. CONVERTS TRANS-2-ENOYL-COA VIA D-3-  
 CC HYDROXYACYL-COA TO 3-KETOACYL-COA.  
 CC -I- PATHWAY: BETA-OXIDATION PATHWAY.  
 CC -I- SUBUNIT: MONOMER (BY SIMILARITY).  
 CC -I- SUBCELLULAR LOCATION: PEROXISOMAL.  
 CC -I- DOMAIN: CONTAINS TWO SDR DOMAINS.  
 CC -I- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES  
 CC (SDR) FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@sdb-sib.ch](mailto:license@sdb-sib.ch)).  
 CC -----  
 DR EMBL: M86456; AAA34779.1; -  
 DR EMBL: X65124; CAA6243.1; -  
 DR EMBL: Z28234; CAA82079.1; -  
 DR PIR: S25322; S25322.  
 DR HSSP: P25529; 1AHH.  
 DR SGD: S0001717; FOX2.  
 DR InterPro: IPR002198; ADH\_SHORT.  
 DR InterPro: IPR002347; Adh\_short\_C2.  
 DR InterPro: IPR002359; MacC\_dehydratase.  
 DR Pfam: PF00106; adh\_short; 2.  
 DR Pfam: PF01575; MacC\_dehydratase; 1.  
 DR PRINTS: PR00080; SDRFAMILY.  
 DR PROSITE: PS00061; ADH\_SHORT; 2.  
 DR PROSITE: PS00342; MICROBODIES\_CTER; 1.  
 DR Fatty acid metabolism: Multifunctional enzyme; Oxidoreductase; NAD;  
 DR Lyase; Isomerase; Peroxisome; Repeat.  
 FT DOMAIN 6 230  
 FT SITE 319 535  
 FT SITE 898 900  
 FT NP\_BIND 13 37  
 FT ACT\_SITE 165 165  
 FT NP\_BIND 326 350  
 FT ACT\_SITE 469 469  
 FT SEQUENCE 900 AA; 98703 MW; 66FFDD049C673788 CRC64;  
 Query Match Score 27; DB 1; Length 900;  
 Best Local Similarity 83.3%; Pred. No. 5e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSLGCS 6  
|:||||  
Db 43 GTLGS 48

```
RESULT 39
AUXL_BOVIN STANDARD; PRT: 910 AA.
ID AUXL_BOVIN
AC Q27974;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_Taxid:9913;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RC TISSUE=Brain;
RA MEDLINE=95220355; PubMed=7705342;
RA Schroeder S., Morris S.A., Knorr R., Plessmann U., Weber K.,
RA Vinh N.G., Ungewickell E.;
RT "Primary structure of the neuronal clathrin-associated protein
RT auxilin and its expression in bacteria.";
RL Eur. J. Biochem. 228:297-304(1995).
CC -1- FUNCTION: BINDS CLATHRIN HEAVY CHAIN AND PROMOTES ITS ASSEMBLY
CC INTO REGULAR CAGES.
CC -1- TISSUE SPECIFICITY: BRAIN.
CC -1- PTM: THE N-TERMINUS IS BLOCKED.
CC -1- PTM: TARGET FOR COAT-ASSOCIATED CASEIN KINASE II IN VITRO.
CC -1- SIMILARITY: CONTAINS 1 TENSIN DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 J DOMAIN.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U09237; AAA79037.1; -
DR InterPro: IPR001623; DnaJ_N.
DR InterPro: IPR000387; TYR_phosphatase.
DR Pfam: PF00226; DnaJ_1.
DR SMART: SM00271; DnaJ_1.
DR PROSITE: PS00383; TYR_PHOSPHATASE_1; UNKNOWN_1.
DR PROSITE: PS00056; TYR_PHOSPHATASE_2; UNKNOWN_1.
DR PROSITE: PS00636; DnaJ_1; FALSE_NEG.
DR PROSITE: PS50076; DnaJ_2; 1.
DR Repeat: Phosphorylation.
FT DOMAIN 33 44 3 X 4 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 33 36 1.
FT REPEAT 37 40 2.
FT REPEAT 41 44 3.
FT DOMAIN 60 344 TENSIN.
FT DOMAIN 406 414 SH3-BINDING (POTENTIAL).
FT DOMAIN 463 757 PRO-RICH.
FT DOMAIN 526 529 POLY-GLY.
FT DOMAIN 846 910 J-DOMAIN.
SO SEQUENCE 910 AA; 99512 MW; BC156DCICF3487FD CRC64;
```

Query Match 90.0%; Score 27; DB 1; Length 910;  
Best Local Similarity 83.3%; Pred. No. 5.1e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSLGCS 6  
|:||||  
Db 680 GTLGS 685

```
RESULT 40
HGBA_HAEDU STANDARD; PRT: 972 AA.
ID HGBA_HAEDU
AC 047952;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE HEMOGLOBIN BINDING PROTEIN PRECURSOR.
GN HGBA.
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_Taxid=730;
RN [1]
RP SEQUENCE FROM N.A.
RC SPRAIN=35000;
RX MEDLINE=95286263; PubMed=7768598;
RA Elkins C., Chen C.J., Thomas C.E.;
RT "Characterization of the hgba locus encoding a hemoglobin receptor
RT from Haemophilus ducreyi.";
RL Infect. Immun. 63:2194-2200(1995).
CC -1- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OF THE HOST AND IS
CC REQUIRED FOR HEME UPTAKE.
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE TONB-DEPENDENT RECEPTOR PROTEIN FAMILY;
CC HEMOGLOBIN/HAPOGLOBIN BINDING PROTEIN SUBFAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U17281; AAA75311.1; -
DR InterPro: IPR000531; TonB_boxC.
DR Pfam: PF00593; TonB_boxC_1.
DR PROSITE: PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
DR PROSITE: PS01156; TONB_DEPENDENT_REC_2; FALSE_NEG.
DR Outer membrane; Transport; TonB box; Multigene family; Signal;
KW Receptor.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 972 HEMOGLOBIN BINDING PROTEIN.
FT SITE 32 39 TONB BOX.
FT SITE 955 972 TONB C-TERMINAL BOX.
SO SEQUENCE 972 AA; 110937 MW; DEDF0F9F3575FD4F CRC64;
```

Query Match 90.0%; Score 27; DB 1; Length 972;  
Best Local Similarity 83.3%; Pred. No. 5.5e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSLGCS 6  
|:||||  
Db 156 GTLGS 161

Search completed: February 4, 2002, 08:01:13  
Job time: 57 sec



Mon Feb 4 08:21:05 2002

us-09-642-660-11.rsp

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 4, 2002, 08:00:52 ; Search time 23.49 Seconds

(Without alignments)  
19.457 Million cell updates/sec

Title: US-09-642-660-11

Perfect score: 50  
Sequence: 1 GSLGGS 6

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database : PIR-68:  
1: p1r1:  
2: p1r2:  
3: p1r3:  
4: p1r4:

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	30	100.0	107	2	G72668 hypothetical prote
2	30	100.0	281	2	S44230 strf protein - Str
3	30	100.0	444	2	S39671 phosphotransferase
4	30	100.0	452	2	F69651 PTS lichenan-speci
5	30	100.0	931	2	S66574 transferrin-bindin
6	30	100.0	935	2	T48778 hypothetical prote
7	30	100.0	936	2	F75622 hypothetical prote
8	30	100.0	1051	2	T51904 hypothetical prote
9	30	100.0	1346	2	A57376 probable regulator
10	30	100.0	1557	2	T02859 probable serine/th
11	28	93.3	166	2	I51242 luteinizing hormon
12	28	93.3	252	2	T48422 hypothetical prote
13	28	93.3	267	2	S44225 strf protein - Str
14	28	93.3	425	2	A38153 paired box protein
15	28	93.3	504	2	T00450 probable monosach
16	28	93.3	562	2	E72608 Mutator-like trans
17	28	93.3	569	2	B84470 Mutator-like trans
18	28	93.3	571	2	C83527 prolyl-tRNA synth
19	28	93.3	572	1	YPEC proline-tRNA lig
20	28	93.3	572	2	C64089 proline-tRNA lig
21	28	93.3	572	2	D85504 proline tRNA synth
22	28	93.3	654	2	T08600 hypothetical prote
23	28	93.3	675	2	T47378 probable transpos
24	28	93.3	729	2	G96559 hypothetical prote
25	28	93.3	749	2	B86403 probable mutator-1
26	28	93.3	758	2	S7855 hypothetical prote
27	28	93.3	851	2	A83484 probable heme uti
28	28	93.3	882	2	T01168 hypothetical prote
29	28	93.3	895	2	T02597 Mutator-like trans

30	28	93.3	940	2	JE0291 FBI9 protein - hum
31	28	93.3	1028	2	E85089 probable transpos
32	28	93.3	1127	2	E86386 probable mutator-1
33	28	93.3	1166	2	T13958 synGAP-D1 protein
34	28	93.3	1249	2	T14270 Ras-GTPase activat
35	28	93.3	1293	2	T14259 Ras GTPase-activat
36	28	93.3	1402	2	S62557 probable calcium-t
37	27	90.0	116	2	S50805 hypothetical prote
38	27	90.0	149	2	S12192 hypothetical prote
39	27	90.0	180	2	G70897 hypothetical prote
40	27	90.0	207	2	JE0045 imidazoleglycerol-
41	27	90.0	231	2	T08181 superoxide dismuta
42	27	90.0	242	2	G82642 conserved hypothet
43	27	90.0	245	2	F84270 ornithine carbamoy
44	27	90.0	267	2	S44228 strf protein - Str
45	27	90.0	277	2	E96769 hypothetical prote
46	27	90.0	280	2	A35872 steroid hormone re
47	27	90.0	281	2	S17776 strf protein - Str
48	27	90.0	283	2	I39625 carbon-monoxide de
49	27	90.0	286	2	B70833 carbon-monoxide de
50	27	90.0	305	2	D96769 hypothetical prote
51	27	90.0	310	2	T20229 hypothetical prote
52	27	90.0	312	2	F70731 hypothetical prote
53	27	90.0	315	2	A35199 heme oxygenase (de
54	27	90.0	325	2	T36851 probable transcrip
55	27	90.0	336	2	F86319 hypothetical prote
56	27	90.0	337	2	D86590 hypothetical prote
57	27	90.0	337	2	G72033 hypothetical prote
58	27	90.0	345	2	A37845 carboxylesterase (
59	27	90.0	349	2	A71212 probable oligopept
60	27	90.0	349	2	H72033 hypothetical prote
61	27	90.0	349	2	E86590 hypothetical prote
62	27	90.0	350	2	I73338 macrophage scaveng
63	27	90.0	350	2	H86371 hypothetical prote
64	27	90.0	353	2	A75027 oligopeptide trans
65	27	90.0	354	2	G75619 probable hemin ABC
66	27	90.0	365	2	F72033 hypothetical prote
67	27	90.0	365	2	B81505 hypothetical prote
68	27	90.0	365	2	C86590 hypothetical prote
69	27	90.0	368	2	H71600 riflin PFB1010W - m
70	27	90.0	378	2	UC5649 N-acetylglucosamin
71	27	90.0	378	2	E82254 N-acetylglucosamin
72	27	90.0	385	2	S46529 polygalacturonase
73	27	90.0	385	2	S46530 polygalacturonase
74	27	90.0	385	2	S46532 polygalacturonase
75	27	90.0	394	2	D70674 probable lipid car
76	27	90.0	396	2	S46533 advanced glycosyla
77	27	90.0	404	1	I61596 macrophage scaveng
78	27	90.0	458	2	B44407 toxin-A - jellyfis
79	27	90.0	463	2	UC7372 probable transpor
80	27	90.0	464	2	D64842 adhesin p1 homolog
81	27	90.0	465	2	S73875 protochlorophyllid
82	27	90.0	466	1	A36904 T-cell surface gly
83	27	90.0	495	1	A26396 histone deacetyl
84	27	90.0	501	2	T05640 usp protein - fru
85	27	90.0	508	2	S11513 steroid hormone re
86	27	90.0	508	2	T13737 probable trypsin V
87	27	90.0	548	2	D82175 hypothetical prote
88	27	90.0	573	2	T10560 transport system P
89	27	90.0	580	1	S01830 hypothetical prote
90	27	90.0	586	2	T17333 cytmH protein - Kle
91	27	90.0	598	2	S63617 hypothetical filam
92	27	90.0	621	2	T48187 intermediate filam
93	27	90.0	643	2	I50539 hypothetical prote
94	27	90.0	656	2	D96831 hypothetical prote
95	27	90.0	713	2	T20090 antrogon receptor
96	27	90.0	716	2	G01627 ferric angubactin
97	27	90.0	726	2	B29928 adhesin p1 precurs
98	27	90.0	737	2	S73792 hypothetical prote
99	27	90.0	746	2	T05899 hypothetical prote
100	27	90.0	769	2	T22256

## ALIGNMENTS

RESULT 1  
G72668  
hypothetical protein APE0773 - Aeropyrum pernix (strain K1)  
C:Species: Aeropyrum pernix  
C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
C:Accession: G72668  
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takai, DNA, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K  
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr  
A:Reference number: A72450; MUID:99310339  
A:Accession: G72668  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-107 <KAM>  
A:Cross-references: DDBJ:AP000060; NID:95104188; PIDN:BAA79751.1; PID:95104436  
A:Experimental source: strain K1  
C:Genetics:  
A:Gene: APE0773  
C:Superfamily: Aeropyrum pernix hypothetical protein APE0773

Query Match 100.0%; Score 30; DB 2; Length 107;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GSIGGS 6  
|||||  
DB 63 GSIGGS 68

RESULT 2  
S44230  
strf protein - Streptomyces glaucoscens  
C:Species: Streptomyces glaucoscens  
C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 10-Dec-1999  
R:Mayer, G.; Pilepersberg, W.  
A:Reference number: S44230  
A:Accession: S44230  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-281 <NAV>  
A:Cross-references: EMBL:X78974; NID:9475235; PIDN:CAA55572.1; PID:9581673  
C:Genetics:  
A:Start codon: GTG  
C:Superfamily: Streptomyces strf protein

Query Match 100.0%; Score 30; DB 2; Length 281;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GSIGGS 6  
|||||  
DB 91 GSIGGS 96

RESULT 3  
S39671  
phosphotransferase system enzyme II homolog ywba - Bacillus subtilis  
N:Alternate names: protein ipa-16d  
C:Species: Bacillus subtilis  
C>Date: 07-Oct-1994 #sequence\_revision 26-May-1995 #text\_change 20-Jun-2000  
C:Accession: S39671; G70050  
R:Glaeher, P.; Kunst, F.; Arnaud, M.; Coudart, M.P.; Gonzales, W.; Hullo, M.F.; Ionescu, M.; Microbiol. 10, 371-384, 1993  
A:Title: Bacillus subtilis genome project: cloning and sequencing of the 97 kb region fl

A:Reference number: S39655; MUID:95020537  
A:Accession: S39671  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-444 <GLA>  
A:Cross-references: EMBL:X73124; NID:9413923; PIDN:CAA51572.1; PID:9413940  
R:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1993  
A:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber  
C: Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;  
Nature 390, 249-256, 1997  
A:Authors: S.D. Emerson, P.T. Ertlan, K.D. Errington, J. Fabbri, C. Ferrari,  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili  
A:Reference number: A69580; MUID:98044033  
A:Accession: G70050  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-444 <KUN>  
A:Cross-references: GB:299123; GB:AL009126; NID:92636240; PIDN:CAB15865.1; PID:926363  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: ywba  
C:Superfamily: phosphotransferase system enzyme II factor II, phosphoenolpyruvate-dep  
C:Keywords: transmembrane protein

Query Match 100.0%; Score 30; DB 2; Length 444;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GSIGGS 6  
|||||  
DB 232 GSIGGS 237

RESULT 4  
F69651  
PTS Iichenan-specific enzyme IIC component IIC - Bacillus subtilis  
N:Alternate names: probable cellobiose phosphotransferase enzyme II'  
C:Species: Bacillus subtilis  
C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Jun-2000  
C:Accession: F69651; S57760  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber  
C: Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;  
A: Enrich, S.D.; Emerson, P.T.; Ertlan, K.D.; Errington, J.; Fabbri, C.; Ferrari,  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi, A.; Gal  
leeh, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M  
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardin  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Ma  
Y, M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Paro, V.; Pohl, T.M.; Portet  
Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scani  
A:Authors: Schleich, S.; Schoeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchida  
T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida  
A:Authors: Yoshikawa, H.F.; Zunstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili  
A:Reference number: A69580; MUID:98044033  
A:Accession: F69651  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-452 <KUN>  
A:Cross-references: GB:299123; GB:AL009126; NID:92636240; PIDN:CAB15884.1; PID:926363  
A:Experimental source: strain 168

R:Glaser, P.; Lubochinsky, B.; Danchin, A.  
submitted to the EMBL Data Library, July 1995  
A:Description: Bacillus subtilis cel operon.  
A:Reference number: S57758

A:Accession: S57760  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-452 <GUA>  
A:Cross-references: EMBL:Z49992; NID:g895746; PIDN:CAA90286.1; PID:g895749  
C:Genetics:  
A:Gene: llec  
A:Start codon: GTG  
C:Superfamily: phosphotransferase system enzyme II factor II, phosphoenolpyruvate-depend

Query Match 100.0%; Score 30; DB 2; Length 452;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GSLSGS 6  
|||||  
Db 232 GSLSGS 237

RESULT 5  
S66574  
transferrin-binding protein 1 precursor - Actinobacillus pleuropneumoniae

C:Species: Actinobacillus pleuropneumoniae  
C:Date: 28-Oct-1998 #sequence-revision 07-Feb-1997 #text-change 26-Aug-1999  
C:Accession: S66574; S53393  
R:Daban, M.; Medrano, A.; Querol, E.  
Biochem. J. 315, 257-264, 1996  
A:Title: Cloning, sequencing and expression of the transferrin-binding protein 1 gene fr  
A:Reference number: S66574; MUID:56207589  
A:Accession: S66574  
A:Molecule type: DNA  
A:Residues: 1-931 <DAB>  
A:Cross-references: EMBL:Z49708; NID:g853708; PIDN:CAA9810.1; PID:g853709  
C:Genetics:  
A:Gene: tbp1  
C:Superfamily: bacterial pathogen transferrin-binding protein; tonB-dependent receptor a  
C:Keywords: Iron transport; membrane protein; metal binding; receptor  
F:1-22/Domain: signal sequence #status predicted <SIG>  
F:23-931/Product: transferrin-binding protein 1 #status predicted <MAT>  
F:67-203/Domain: tonB-dependent receptor amino-terminal homology <TNB>  
F:600-931/Domain: tonB-dependent receptor carboxyl-terminal homology <TNC>

Query Match 100.0%; Score 30; DB 2; Length 931;  
Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GSLSGS 6  
|||||  
Db 151 GSLSGS 156

RESULT 6

T48778  
hypothetical protein J3E11.260 [Imported] - Neurospora crassa  
C:Species: Neurospora crassa  
C:Date: 05-May-2000 #sequence-revision 05-May-2000 #text-change 05-May-2000  
C:Accession: T48778  
R:Schulte, U.; Algn, V.; Hohelsel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,  
submitted to the Protein Sequence Database, April 2000  
A:Reference number: Z24541  
A:Accession: T48778  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-935 <SCH>  
A:Cross-references: EMBL:AL353820; GSPDB:GN00112; NCSP:J3E11.260  
A:Experimental source: cosmid contig J3E11; strain 74  
C:Genetics:

A:Gene: NCSP:J3E11.260  
A:Map position: 2

Query Match 100.0%; Score 30; DB 2; Length 935;  
Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GSLSGS 6  
|||||  
Db 700 GSLSGS 705

RESULT 7

F75622  
hypothetical protein - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C:Date: 03-Dec-1999 #sequence-revision 03-Dec-1999 #text-change 31-Mar-2000  
C:Accession: F75622  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896  
A:Accession: F75622  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-936 <WHI>  
A:Cross-references: GB:AE001826; NID:g6460827; PIDN:AF12631.1; PID:g6460927; TIGR:DR  
C:Experimental source: strain R1  
C:Genetics:  
A:Gene: DRB0041  
A:Map position: megaplasmid  
A:Genome: plasmid  
A:Note: plasmid MP1

Query Match 100.0%; Score 30; DB 2; Length 936;  
Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GSLSGS 6  
|||||  
Db 800 GSLSGS 805

RESULT 8

T51904  
hypothetical protein B2311.160 [Imported] - Neurospora crassa  
C:Species: Neurospora crassa  
C:Date: 20-Oct-2000 #sequence-revision 20-Oct-2000 #text-change 20-Oct-2000  
C:Accession: T51904  
R:Schulte, U.; Algn, V.; Hohelsel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatu  
submitted to the Protein Sequence Database, August 2000  
A:Reference number: Z25858  
A:Accession: T51904  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1051 <SCH>  
A:Cross-references: EMBL:AL391572; GSPDB:GN00116; NCSP:B2311.160  
A:Experimental source: BAC clone B2311; strain OR74A  
C:Genetics:  
A:Gene: NCSP:B2311.160  
A:Map position: 6

Query Match 100.0%; Score 30; DB 2; Length 1051;  
Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GSLSGS 6  
|||||

Db 828 GSLGGS 833

## RESULT 9

A57376

Probable regulatory protein 322 - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 08-Feb-1996 #sequence\_revision 08-Feb-1996 #text\_change 08-Feb-1996

C:Accession: A57376

R:Lin, X.; Nelson, P.J.; Frankfort, B.; Tomblar, E.; Johnson, R.; Gelman, I.H.

Mol. Cell. Biol. 15, 2754-2762, 1995

A:Title: Isolation and characterization of a novel mitogenic regulatory gene, 322, which

A:Reference number: A57376; MUID:95257957

A:Accession: A57376

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-1346 &lt;LIN&gt;

A:Cross-references: GB:U23146

C:Genetics:

A:Gene: 322

Query Match

Best Local Similarity 100.0%; Score 30; DB 2; Length 1346;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLGGS 6

Db 960 GSLGGS 965

## RESULT 10

T02859

Probable serine/threonine kinase STPK1 [Imported] - Leishmania major (strain Friedlin)

C:Species: Leishmania major

C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 19-May-2000

C:Accession: G81463; T02859

R:Wyer, P.J.; Audleman, L.; deVos, T.; Hixson, G.; Kiser, P.; Lemley, C.; Magness, C.;

Proc. Natl. Acad. Sci. U.S.A. 96, 2902-2906, 1999

A:Title: Leishmania major Friedlin chromosome 1 has an unusual distribution of protein-c

A:Reference number: A81455; MUID:99178987

A:Accession: G81463

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1557 &lt;PYL&gt;

A:Cross-references: GB:AE001274; NID:93264850; PIDN:AAC24682.1; PID:93002481; GSPDB:GNOC

A:Experimental source: strain MHOM/IL/81/Friedlin

C:Genetics:

A:Gene: STPK1

A:Map position: 1

C:Superfamily: Leishmania major probable membrane protein STPK1

Query Match

Best Local Similarity 100.0%; Score 30; DB 2; Length 1557;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLGGS 6

Db 158 GSLGGS 163

## RESULT 11

I51242

Inteinizing hormone beta-subunit - quail

C:Species: Coturnix coturnix (quail)

C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 16-Jul-1999

C:Accession: I51242

R:Ando, H.; Ishii, S.

Gen. Comp. Endocrinol. 93, 357-368, 1994

A:Title: Molecular cloning of complementary deoxyribonucleic acids for the pituitary gly

A:Reference number: I51241; MUID:94252550

A:Accession: I51242

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-166 &lt;AND&gt;

A:Cross-references: GB:S70834; NID:95469921; PIDN:AB30867.1; PID:95469922

C:Superfamily: pituitary glycoprotein hormone beta chain

Query Match

Best Local Similarity 93.3%; Score 28; DB 2; Length 166;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLGGS 6

Db 46 GSMGGS 51

## RESULT 12

T48422

Hypothetical protein F8F6.160 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000

C:Accession: T48422

R:Bevan, M.; Peters, S.A.; van Staveren, M.; Dirke, W.; Stiekema, W.; Bancroft, I.;

Submitted to the Protein Sequence Database, March 2000

A:Reference number: 224488

A:Accession: T48422

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-252 &lt;BEV&gt;

A:Cross-references: EMBL:AL162873

A:Experimental source: cultivar Columbia; BAC clone F8F6

C:Genetics:

A:Map position: 5

A:Note: F8F6.160

Query Match

Best Local Similarity 93.3%; Score 28; DB 2; Length 252;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLGGS 6

Db 154 GSIGGS 159

## RESULT 13

S44225

strf protein - Streptomyces bluenis (fragment)

C:Species: Streptomyces bluenis

C:Date: 19-Mar-1997 #sequence\_revision 18-Jul-1997 #text\_change 11-Jan-2000

C:Accession: S44225

R:Mayer, G.; Piepersberg, W.

Submitted to the EMBL Data Library, April 1994

A:Reference number: S44224

A:Accession: S44225

A:Molecule type: DNA

A:Residues: 1-267 &lt;MAY&gt;

A:Cross-references: EMBL:X78972; NID:9475227; PIDN:CAA55568.1; PID:9581604

C:Genetics:

A:Start codon: GTG

C:Superfamily: Streptomyces strf protein

Query Match

Best Local Similarity 93.3%; Score 28; DB 2; Length 267;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLGGS 6

Db 91 GSIGGS 96

RESULT 14  
A38153  
paired box protein fox neuro - fruit fly (*Drosophila melanogaster*)  
C:Species: *Drosophila melanogaster*  
C>Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 21-Jul-2000  
C:Accession: A38153; S06951  
R:Damdy-Chaudiere, C.; Jamet, E.; Burri, M.; Bopp, D.; Basler, K.; Hafen, E.; Dumont, N.  
Cell 69, 159-172, 1992  
A>Title: The paired box gene fox neuro: a determinant of poly-innervated sense organs in  
A:Reference number: A38153; MIMD:92208941  
A:Accession: A38153  
A:Molecule type: DNA  
A:Residues: 1-425 <DAM>  
A:Cross-references: GB:W86927; NID:q158155; PID:q158156  
A>Note: sequence extracted from NCBI backbone (NCBIN:90907, NCBIN:90912, NCBIN:90913)  
R:Bopp, D.; Jamet, E.; Baumgartner, S.; Burri, M.; Noll, M.  
EMBO J. 8, 3447-3457, 1989  
A>Title: Isolation of two tissue-specific *Drosophila* paired box genes, Fox meso and Fox  
A:Reference number: S06950; MIMD:90059940  
A:Accession: S06951  
A:Molecule type: DNA  
A:Residues: 1-135 <BOP>  
A:Cross-references: EMBL:X58917; NID:q8360; PIDN:CAA41721.1; PID:q1405550  
C:Genetics:  
A:Gene: FlyBase:Box-n  
A:Cross-references: FlyBase:Fgn0003130  
C:Superfamily: paired box transcription factor Pax-8; paired box homology  
F:5-130/Domain: paired box homology <PBH>

Query Match  
Best Local Similarity 93.3%; Score 28; DB 2; Length 425;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSIGS 6  
DB 70 GSIGS 75

RESULT 15  
T00450  
probable monosaccharide transport protein T14N5.7 - *Arabidopsis thaliana*  
C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
C>Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 24-Sep-1999  
C:Accession: T00450  
R:Federspiel, N.A.; Palm, C.J.; Conway, A.B.; Kurtz, D.B.; Conway, A.R.; Au, M.; Araujo,  
submitted to the EMBL Data Library, September 1998  
A:Reference number: 214152  
A:Accession: T00450  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-504 <FED>  
A:Cross-references: EMBL:AC004260; NID:q3176694; PIDN:AAC34349.1; PID:q3540199  
C:Genetics:  
A:Map position: 1  
A:introns: 48/1; 155/3; 365/3  
A>Note: T14N5.7  
C:Superfamily: glucose transport protein  
C:Keywords: sugar transport; transmembrane protein

Query Match  
Best Local Similarity 93.3%; Score 28; DB 2; Length 504;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSIGS 6  
DB 33 GSIGS 38

RESULT 16  
E72608

probable hyub APE1331 - *Aeropyrum pernix* (strain K1)  
C:Species: *Aeropyrum pernix*  
C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
C:Accession: E72608  
R:Kawabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Ta  
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.  
DNA Res. 6, 83-101, 1999  
A>Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero  
A:Reference number: A72450; MIMD:99310339  
A:Accession: A72450  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-562 <KAW>  
A:Cross-references: DDBJ:AP000061; NID:q5104821; PIDN:BAAB0323.1; PID:q5105009  
A:Experimental source: strain K1  
C:Genetics:  
A:Gene: APE1331  
C:Superfamily: N-methylhydantoinase (ATP-hydrolyzing)

Query Match  
Best Local Similarity 93.3%; Score 28; DB 2; Length 562;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSIGS 6  
DB 136 GSIGS 141

RESULT 17  
B84470  
Mutator-like transposase [imported] - *Arabidopsis thaliana*  
C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: B84470  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Beutlo, M.I.; Town, C.D.; Fujii, C.Y.  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter  
Nature 402, 761-768, 1999  
A>Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.  
A:Reference number: A84420; MIMD:20083487  
A:Accession: B84470  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-569 <STO>  
A:Cross-references: GB:AE002093; NID:q4581167; PIDN:AAD24650.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: AT290570  
A:Map position: 2

Query Match  
Best Local Similarity 93.3%; Score 28; DB 2; Length 569;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSIGS 6  
DB 165 GSIGS 170

RESULT 18  
CB3527  
prolyl-tRNA synthetase PA0956 [imported] - *Pseudomonas aeruginosa* (strain PA01)  
C:Species: *Pseudomonas aeruginosa*  
C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: CB3527  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; L  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A>Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pa  
A:Reference number: A82950; MIMD:20437337  
A:Accession: CB3527

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-571 <STO>  
A:Cross-references: GB:AE004529; GB:AE004091; NID:99946851; PIDN:AG04345.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: pros: PA0956  
C:Superfamily: proline--tRNA ligase

Query Match 93.3%; Score 28; DB 2; Length 571;  
Best Local Similarity 83.3%; Pred. No. 3.8e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSIGS 6  
DB 200 GSIGS 205

RESULT 19  
typeC  
Proline--tRNA ligase (EC 6.1.1.15) - Escherichia coli  
N:Alternate names: global RNA synthesis factor; prolyl-tRNA synthetase  
C:Species: Escherichia coli  
C:Date: 30-Sep-1992 #sequence,revision 31-Oct-1997 #text\_change 24-Sep-1999  
C:Accession: B64744; J00110; J00061  
R:Baltner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co  
A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617  
A:Accession: B64744

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA  
A:Residues: 1-572 <BLAT>  
A:Cross-references: GB:AE000128; GB:000096; NID:91786393; PIDN:AACT3305.1; PID:91786392;  
A:Experimental source: strain K-12, substrain MG1655  
R:Erland, G.; Delarue, M.; Poch, O.; Gangloff, J.; Moras, D.  
Nature 347, 203-206, 1990  
A:Title: Partition of tRNA synthetases into two classes based on mutually exclusive sets  
A:Reference number: J00110; MUID:90370122  
A:Accession: J00110

A:Molecule type: DNA  
A:Residues: 1-25, 'IV', '28-204', 'Q', '206-572' <ERI>  
A:Cross-references: GB:M97858; NID:9147361; PIDN:AAA24420.1; PID:9147362  
A:Note: the sequence from Fig. 2 is inconsistent with that from Fig. 1 in having 205-ser  
R:Zhou, Z.; Syvanen, M.  
J. Bacteriol. 172, 281-286, 1990  
A:Title: Identification and sequence of the drpA gene from Escherichia coli.  
A:Reference number: J00061; MUID:90094229  
A:Accession: J00061

A:Molecule type: DNA  
A:Residues: 1-126, 'HV', '129-204', 'RPLTNSRCWR', '217-515', 'RA', <ZHO>  
A:Cross-references: GB:M32357; NID:9145802; PIDN:AAA23710.1; PID:9145803  
A:Note: the authors identified this protein as global RNA synthesis factor (encoded by c  
C:Genetics:  
A:Gene: pros: drpA  
A:Map position: 5 min  
C:Superfamily: proline--tRNA ligase  
C:Keywords: aminoacyl-tRNA synthetase; ATP; homodimer; ligase; protein biosynthesis

Query Match 93.3%; Score 28; DB 1; Length 572;  
Best Local Similarity 83.3%; Pred. No. 3.8e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSIGS 6  
DB 200 GSIGS 205

RESULT 20  
CG4089

proline--tRNA ligase (EC 6.1.1.15) - Haemophilus influenzae (strain Rd KW20)  
N:Alternate names: prolyl-tRNA synthetase  
C:Species: Haemophilus influenzae  
C:Date: 18-Aug-1995 #sequence,revision 18-Aug-1995 #text\_change 16-Jul-1999  
C:Accession: C64089  
R:Flischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage  
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman  
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.  
Science 269, 496-512, 1995  
A:Authors: Guehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente  
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A:Reference number: A64000; MUID:95350630  
A:Accession: C64089

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA  
A:Residues: 1-572 <TRIG>  
A:Cross-references: GB:U32756; GB:LA2023; NID:91573729; PIDN:AACT2388.1; PID:91573733  
C:Genetics:  
A:Gene: pros  
C:Superfamily: proline--tRNA ligase  
C:Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 93.3%; Score 28; DB 2; Length 572;  
Best Local Similarity 83.3%; Pred. No. 3.8e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSIGS 6  
DB 200 GSIGS 205

RESULT 21  
D85504  
Proline tRNA synthetase [imported] - Escherichia coli (strain O157:H7)  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence,revision 16-Feb-2001 #text\_change 31-Mar-2001  
C:Accession: D85504  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May  
; Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206531  
A:Accession: D85504

A:Molecule type: DNA  
A:Residues: 1-572 <STO>  
A:Cross-references: GB:AE005174; NID:912512925; PIDN:AAAG54496.1; GSPDB:GN00145; UMGP:  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: pros  
C:Superfamily: proline--tRNA ligase

Query Match 93.3%; Score 28; DB 2; Length 572;  
Best Local Similarity 83.3%; Pred. No. 3.8e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSIGS 6  
DB 200 GSIGS 205

RESULT 22  
T08600  
hypothetical protein Opa - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C:Date: 11-Jun-1999 #sequence,revision 11-Jun-1999 #text\_change 17-Nov-2000  
C:Accession: T08600  
R:Wagner, R.A.; Friedman, T.B.; Smith, B.; Wallrath, L.L.; Johnson, S.; Burnett, J.B.  
submitted to the EMBL Data Library, August 1997  
A:Reference number: Z16450  
A:Accession: T08600



A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-654 <MAG>  
A:Cross-references: EMBL:AF018078; NID:g2394383; PID:g2394384  
A:Experimental source: strain Oregon-R  
C:Genetics:  
A:Gene: Opa  
A:Cross-references: FlyBase:FBgn0020880  
A:Map position: 2

Query Match 93.3%; Score 28; DB 2; Length 654;  
Best Local Similarity 83.3%; Pred. No. 4.4e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSIGS 6  
Db 387 GSIGS 392

## RESULT 23

T47378  
probable transposase - Arabidopsis thaliana  
N:Alternate names: protein T5C2.60  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
C:Accession: T47378  
R:Dermlater, B.; Ottenwaelder, B.; Duchemin, D.; Zeidler, K.; Mewes, H.W.; Lemcke, K.; M  
submitted to the Protein Sequence Database, March 2000  
A:Reference number: 224463  
A:Accession: T47378  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-675 <OBES>  
A:Cross-references: EMBL:AL138664  
A:Experimental source: cultivar Columbia; BAC clone T5C2  
C:Genetics:  
A:Map position: 3  
A:Note: T5C2.60

Query Match 93.3%; Score 28; DB 2; Length 675;  
Best Local Similarity 83.3%; Pred. No. 4.5e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSIGS 6  
Db 180 GSIGS 185

## RESULT 24

G96559  
hypothetical protein F5F19.7 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: G96559  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Hulzar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maitl, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: AB6141; MUID:21016719  
A:Accession: G96559  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-729 <STO>  
A:Cross-references: GB:AE005173; NID:g4220448; PIDN:ADJ12675.1; GSPDB:GN00141  
C:Genetics:

A:Gene: F5F19.7  
A:Map position: 1

Query Match 93.3%; Score 28; DB 2; Length 729;  
Best Local Similarity 83.3%; Pred. No. 4.9e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSIGS 6  
Db 163 GSIGS 168

## RESULT 25

B86403  
probable mutator-like transposase [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: B86403  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,  
ansen, N.F.; Hughes, B.; Hulzar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maitl, R.; Marzia  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tall  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: AB6141; MUID:21016719  
A:Accession: B86403  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-749 <STO>  
A:Cross-references: GB:AE005172; NID:g10998920; PIDN:AGC26060.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1

Query Match 93.3%; Score 28; DB 2; Length 749;  
Best Local Similarity 83.3%; Pred. No. 5e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSIGS 6  
Db 180 GSIGS 185

## RESULT 26

S37855  
hypothetical protein YKL034w precursor - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein YKL247  
C:Species: Saccharomyces cerevisiae  
C:Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 02-Sep-2000  
C:Accession: S37855; S41670; S36853  
R:Purnelle, B.; Skala, J.; van Dyck, L.; Tettelin, H.; Goffeau, A.  
submitted to the Protein Sequence Database, March 1994  
A:Reference number: S37851  
A:Accession: S37855  
A:Molecule type: DNA  
A:Residues: 1-758 <PUR>  
A:Cross-references: EMBL:Z28034; NID:g486043; PIDN:CAAB1869.1; PID:g486044; MIPS:YKLO  
A:Experimental source: strain S288C  
R:Purnelle, B.; Skala, J.; van Dyck, L.; Goffeau, A.  
Yeast 10, 125-130, 1994  
A:Title: Analysis of an 11.7 kb DNA fragment of chromosome XI reveals a new tRNA gene  
guilator ABF2.  
A:Reference number: S41667; MUID:94262309  
A:Accession: S41670  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-758 <PUD>  
A:Cross-references: EMBL:X71622

A:Experimental source: strain S286C  
R:Purnelle, B: Skala, J: van Dyck, L: Goffeau, A.  
A:Title: The sequence of a 12 kb fragment on the left arm of yeast chromosome XI reveals  
se from potato.  
A:Reference number: S30007; MUID:91127731  
A:Accession: S36853  
A:Status: translation not shown  
A:Molecule type: DNA  
A:Residues: 1-570 <PU3>  
A:Cross-references: EMBL:X69584; NID:94789; PIDN:CAA9298.1; PID:9871537  
A:Experimental source: strain S286C  
C:Genetics:  
A:Map position: 11L  
C:Superfamily: Saccharomyces cerevisiae hypothetical protein YKL034w; RING finger homolo  
C:Keywords: transmembrane protein  
F:1-26/Domain: signal sequence #status predicted <SIG>  
F:27-758/Product: hypothetical protein YKL034w #status predicted <MNT>  
F:400-416/Domain: transmembrane #status predicted <TM1>  
F:440-456/Domain: transmembrane #status predicted <TM2>  
F:461-477/Domain: transmembrane #status predicted <TM3>  
F:528-544/Domain: transmembrane #status predicted <TM4>  
F:607-623/Domain: transmembrane #status predicted <TM5>  
F:638-654/Domain: transmembrane #status predicted <TM6>  
F:695-757/Domain: RING finger homology <RRN>

Query Match 93.3%; Score 28; DB 2; Length 758;  
Best Local Similarity 83.3%; Pred. No. 5.1e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLOGS 6  
DB 522 GSLOGS 527

RESULT 27  
A83484  
probable heme utilization protein precursor PA1302 [imported] - Pseudomonas aeruginosa  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: A83484  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B  
achman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardi, K.; Lim,  
; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: A82950; MUID:20437337  
A:Accession: A83484  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-851 <STO>  
A:Cross-references: GB:AE004559; GB:AE004091; NID:95947228; PIDN:AG04691.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA1302

Query Match 93.3%; Score 28; DB 2; Length 851;  
Best Local Similarity 83.3%; Pred. No. 5.7e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLOGS 6  
DB 668 GSLOGS 673

RESULT 28  
T01168  
hypothetical protein F7N22.13 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 18-Aug-2000  
C:Accession: T01168

R:Dante, M.  
Submitted to the EMBL Data Library, April 1998  
A:Description: The sequence of A. thaliana F7N22.  
A:Reference number: Z14250  
A:Accession: T01168  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-882 <DAN>  
A:Cross-references: EMBL:AF058825; NID:93047060; PID:93047071  
A:Experimental source: cultivar Columbia  
C:Genetics:  
A:Map position: 4  
A:Introns: 60/71; 852/3  
A:Note: F7N22.13  
C:Superfamily: Arabidopsis thaliana hypothetical protein F26C24.7

Query Match 93.3%; Score 28; DB 2; Length 882;  
Best Local Similarity 83.3%; Pred. No. 5.9e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLOGS 6  
DB 128 GSLOGS 133

RESULT 29  
T02597  
Mutator-like transposase [imported] - Arabidopsis thaliana  
N:Alternate names: hypothetical protein F26C24.7  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 16-Feb-2001  
C:Accession: T02597; E84521  
R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Syke  
submitted to the EMBL Data Library, June 1998  
A:Description: Arabidopsis thaliana chromosome II BAC F26C24 genomic sequence.  
A:Reference number: Z14680  
A:Accession: T02597  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-895 <ROU>  
A:Cross-references: EMBL:AC004705; NID:93252804; PID:93252817  
A:Experimental source: cultivar Columbia  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Bentto, M.I.; Town, C.D.; Fujii, C.Y  
M.; Koo, H.; Moffat, K.S.; Cronlin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487  
A:Accession: E84521  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-895 <STO>  
A:Cross-references: GB:AE002093; NID:93252817; PIDN:AMC24187.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g14790; F26C24.7  
A:Map position: 2  
A:Introns: 70/72; 151/71; 292/3; 420/3  
C:Superfamily: Arabidopsis thaliana hypothetical protein F26C24.7

Query Match 93.3%; Score 28; DB 2; Length 895;  
Best Local Similarity 83.3%; Pred. No. 6e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLOGS 6  
DB 129 GSLOGS 134

RESULT 30  
JE0291  
FB19 protein - human

C:Species: Homo sapiens (man)  
C:Date: 05-Dec-1998 #sequence\_revision 05-Dec-1998 #text\_change 21-Jul-2000  
C:Accession: JE0291  
R:Torcaro, A.; Grifa, A.; Carrella, M.; Rommens, J.M.; Valentino, M.A.; Roetto, A.; Zelanti  
Biochem. Biophys. Res. Commun. 250, 555-557, 1998  
A:Title: Cloning of a new gene (FBI9) within HLA class I region.  
A:Reference number: JE0291; MUID:99003493  
A:Accession: JE0291  
A:Molecule type: mRNA  
A:Residues: 1-940 <TOT>  
A:Cross-references: GB:Y13247; NID:g2117158; PIDN:CAAF3697.1; PID:g2117159  
C:Genetics:  
A:Gene: FBI9  
A:Map position: 6p21.3

Query Match 93.3%; Score 28; DB 2; Length 940;  
Best Local Similarity 83.3%; Pred. No. 6.3e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GSIGS 6  
DB 837 GSMGS 842

RESULT 31  
E83089  
probable transposon protein [Imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001  
C:Accession: E85089  
R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring  
Nature 402, 769-777, 1999  
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
A:Reference number: A85001; MUID:20083488  
A:Accession: E85089  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-1028 <STO>  
A:Cross-references: GB:NC\_001268; NID:g7267531; PIDN:CAB78013.1; GSPDB:GN00140  
C:Genetics:  
A:Gene: AT4g08890  
A:Map position: 4

Query Match 93.3%; Score 28; DB 2; Length 1028;  
Best Local Similarity 83.3%; Pred. No. 6.9e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GSIGS 6  
DB 128 GSIGS 133

RESULT 32  
E86386  
probable mutator-like transposase [Imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: E86386  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chlin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
Nansen, N.F.; Hughes, B.; Hulzar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lucero, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: E86386  
A:Status: preliminary

A:Molecule type: DNA  
A:Residues: 1-1127 <STO>  
A:Cross-references: GB:AE005172; NID:g11560180; PIDN:AAG38122.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1

Query Match 93.3%; Score 28; DB 2; Length 1127;  
Best Local Similarity 83.3%; Pred. No. 7.6e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GSIGS 6  
DB 180 GSIGS 185

RESULT 33  
T13958  
SYNGAP-b1 protein - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C:Accession: T13958  
R:Suzuki, T.  
submitted to the EMBL Data Library, August 1998  
A:Description: SYNGAP-b1  
A:Reference number: Z17834  
A:Accession: T13958  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1166 <SVU>  
A:Cross-references: EMBL:AB016962; NID:d1261311; PID:d1038706; PIDN:CAB19493.1  
C:Genetics:  
A:Experimental source: strain Sprague Dawley  
A:Gene: syngap-b1

Query Match 93.3%; Score 28; DB 2; Length 1166;  
Best Local Similarity 83.3%; Pred. No. 7.9e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GSIGS 6  
DB 996 GSIGS 1001

RESULT 34  
T14270  
Ras-GTPase activating protein SYNGAP-b - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C:Accession: T14270  
R:Kim, J.H.; Liao, D.; Lau, L.F.; Haganir, R.L.  
Neuron 20, 683-691, 1998  
A:Title: SYNGAP: a synaptic RasGAP that associates with the PSD-95/SAP90 protein fam1  
A:Reference number: T17950; MUID:98240917  
A:Accession: T14270  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1249 <KIM>  
A:Cross-references: EMBL:AF058790; NID:g3722228; PID:g3722229; PIDN:AAC63511.1

Query Match 93.3%; Score 28; DB 2; Length 1249;  
Best Local Similarity 83.3%; Pred. No. 8.5e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSIGS 6  
DB 1054 GSIGS 1059

RESULT 35  
T14259

ras GTPase-activating protein, synaptic - rat  
 N:Alternate names: protein SynGAP  
 C:Species: Rattus norvegicus (Norway rat)  
 C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
 C:Accession: T14259  
 R:Chen, H.-J.; Rojas-Soto, M.; Kennedy, M.B.  
 Submitted to the EMBL Data Library, February 1998  
 A:Description: A synaptic Ras GTPase-activating protein (p135 SynGAP) inhibited by Cam K  
 A:Reference number: 217942  
 A:Accession: T14259  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1293 <CHE>  
 A:Cross-references: EMBL:AF048976; NID:g2935447; PID:g2935448; PIDN:MAC08071.1  
 A:Experimental source: strain Sprague-Dawley

Query Match 93.3% Score 28; DB 2; Length 1293;  
 Best Local Similarity 83.3% Pred. No. 8.8e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSIGGS 6  
 ||:||||  
 DB 1098 GSIGGS 1103

RESULT 36  
 S62557  
 Probable calcium-transporting ATPase - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C>Date: 16-May-1996 #sequence\_revision 13-Mar-1997 #text\_change 21-Jan-2000  
 C:Accession: T38339; S62557  
 R:Odell, C.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.  
 Submitted to the EMBL Data Library, November 1995  
 A:Reference number: 221766  
 A:Accession: T38339  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1402 <OD2>  
 A:Cross-references: EMBL:Z67757; NID:g1061288; PIDN:CAA91777.1; PID:g1061300; GSPDB:GN0C  
 A:Experimental source: strain 972h-; cosmid G24B11  
 C:Genetics:  
 A:Gene: SPAC24B11.12c  
 A:Map position: 1L  
 A:Insertions: 124/3

Query Match 93.3% Score 28; DB 2; Length 1402;  
 Best Local Similarity 83.3% Pred. No. 9.5e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSIGGS 6  
 ||:||||  
 DB 490 GSIGGS 495

RESULT 37  
 S50805  
 Hypothetical protein YJL067w - yeast (Saccharomyces cerevisiae)  
 N:Alternate names: hypothetical protein HRA116; hypothetical protein J1107  
 C:Species: Saccharomyces cerevisiae  
 C>Date: 13-Jan-1995 #sequence\_revision 08-Sep-1995 #text\_change 29-Oct-1999  
 C:Accession: S50805; S47124; S56841  
 R:Vandenbol, M.; Durand, P.; Dion, C.; Portetelle, D.; Hilger, F.  
 Yeast 11, 57-60, 1995  
 A:Title: Sequence of a 17.1 kb DNA fragment from chromosome X of Saccharomyces cerevisiae  
 A:Reference number: S50798; MUID:95282514  
 A:Accession: S50805  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-116 <YAN>  
 A:Cross-references: EMBL:Z34288; NID:g498992; PIDN:CAA84056.1; PID:g499000  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1994

R:Vandenbol, M.; Durand, P.; Dion, C.; Portetelle, D.; Hilger, F.  
 Submitted to the EMBL Data Library, June 1994  
 A:Description: Sequence analysis of a 17.1 kb DNA fragment from chromosome X of Sacch  
 A:Reference number: S47117  
 A:Accession: S47124  
 A:Molecule type: DNA  
 A:Residues: 1-116 <YAN>  
 A:Cross-references: EMBL:Z34288; NID:g498992; PIDN:CAA84056.1; PID:g499000  
 R:Vandenbol, M.; Durand, P.; Portetelle, D.; Hilger, F.  
 Submitted to the Protein Sequence Database, September 1995  
 A:Reference number: S56835  
 A:Accession: S56841  
 A:Molecule type: DNA  
 A:Residues: 1-116 <POB>  
 A:Cross-references: EMBL:Z49342; NID:g1008215; PIDN:CAA89357.1; PID:g1008216; MIPS:YJ  
 C:Genetics:  
 A:Map position: 10L  
 C:Superfamily: Saccharomyces hypothetical protein YJL067w

Query Match 90.0% Score 27; DB 2; Length 116;  
 Best Local Similarity 83.3% Pred. No. 1.2e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSIGGS 6  
 ||:||||  
 DB 27 GSIGGS 32

RESULT 38  
 S12192  
 Hypothetical protein 5 - Thibacillus ferrooxidans plasmid pTF1  
 C:Species: Thibacillus ferrooxidans  
 C>Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 25-Apr-1997  
 C:Accession: S12192  
 R:Drolet, M.; Zanga, P.; Lau, P.C.K.  
 Mol. Microbiol. 4, 1381-1391, 1990  
 A:Title: The mobilization and origin of transfer regions of a Thibacillus ferrooxidans  
 A:Reference number: S12188; MUID:91125140  
 A:Accession: S12192  
 A:Status: preliminary; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-149 <DRO>  
 A:Cross-references: EMBL:X52699  
 C:Genetics:  
 A:Genome: plasmid pTF1

Query Match 90.0% Score 27; DB 2; Length 149;  
 Best Local Similarity 83.3% Pred. No. 1.6e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSIGGS 6  
 ||:||||  
 DB 92 GSIGGS 97

RESULT 39  
 G70897  
 Hypothetical protein Ry1105 - Mycobacterium tuberculosis (strain H37RV)  
 C:Species: Mycobacterium tuberculosis  
 C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
 C:Accession: G70897  
 R:Conor, R.; Davies, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
 Rajandream, M.A.; Rogers, J.; Sultson, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 Nature 393, 537-544, 1998  
 A:Authors: Squires, R.; Sultson, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A:Reference number: A70500; MUID:98295987  
 A:Accession: G70897  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA

A:Residues: 1-180 <COL>  
A:Cross-references: GB:AL021897; GB:AL123456; NID:g3256022; PIDN:CAA17221.1; PID:e125198  
A:Experimental source: strain H37Rv  
C:Genetics:  
A:Gene: RV1105

Query Match 90.0%; Score 27; DB 2; Length 180;  
Best Local Similarity 83.3%; Pred. No. 1.9e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GSUGS 6  
|:||||  
Db 98 GNLGGS 103

RESULT 40  
JE0045 Imidazoleglycerol-phosphate dehydratase (EC 4.2.1.19) - Azospirillum brasilense  
C:Species: Azospirillum brasilense  
C>Date: 07-Jun-1990 #sequence.revision 07-Jun-1990 #text\_change 31-Mar-2000  
C:Accession: JE0045; S04399; S16798  
R:Fani, R.; Bazzicalupo, M.; Damiani, G.; Bianchi, A.; Schipani, C.; Sgarbetta, V.; Pol  
Mol. Gen. Genet. 216, 224-229, 1989  
A>Title: Cloning of histidine genes of Azospirillum brasilense: organization of the ABPH  
A:Reference number: JE0045; MID:89313660  
A:Accession: JE0045  
A:Molecule type: DNA  
A:Residues: 1-207 <FAN>  
A:Cross-references: EMBL:X61207; NID:g38669; PIDN:CAA43515.1; PID:g38670  
C:Genetics:  
A:Gene: hlsB  
C:Superfamily: Imidazoleglycerol-phosphate dehydratase; imidazoleglycerol-phosphate dehy  
C:Keywords: carbon-oxygen lyase; histidine biosynthesis; hydro-lyase  
F:36-203/Domain: imidazoleglycerol-phosphate dehydratase homology <IIP>

Query Match 90.0%; Score 27; DB 2; Length 207;  
Best Local Similarity 83.3%; Pred. No. 2.2e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GSUGS 6  
|:||||  
Db 201 GTUGGS 206

Search completed: February 4, 2002, 08:00:55  
Job time: 39 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 4, 2002, 08:02:21 ; Search time 35.81 Seconds  
(Without alignments)  
24.508 Million cell updates/sec

Title: US-09-642-660-11  
Perfect score: 30  
Sequence: 1 GSIGGS 6

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database :

SP\_mammal:\*  
1: sp\_bacteria:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	30	100.0	107	1	Q9YDZ6
2	30	100.0	175	5	Q9VBD9
3	30	100.0	203	12	Q75013
4	30	100.0	209	12	Q75018
5	30	100.0	277	5	Q54259
6	30	100.0	281	2	Q54259
7	30	100.0	286	5	Q9VBE0
8	30	100.0	371	2	Q9X6F4
9	30	100.0	590	4	Q9B0G1
10	30	100.0	750	10	Q9FGC2
11	30	100.0	763	5	Q9U193
12	30	100.0	773	10	Q9AV46
13	30	100.0	802	2	Q9XG84
14	30	100.0	898	2	Q44125
15	30	100.0	931	2	Q44158
16	30	100.0	931	2	Q44127
17	30	100.0	931	2	Q44166
18	30	100.0	935	3	Q9P6Z0
19	30	100.0	936	2	Q9R2S3

20	30	100.0	993	2	Q9K1V0	Q9K1V0 haemophilus
21	30	100.0	999	2	Q87296	Q87296 haemophilus
22	30	100.0	1013	2	Q9K1V2	Q9K1V2 haemophilus
23	30	100.0	1081	5	Q9VGA4	Q9VGA4 drosophila
24	30	100.0	1340	5	Q9VHR2	Q9VHR2 drosophila
25	30	100.0	1557	5	Q60976	Q60976 leishmania
26	30	100.0	1596	11	Q62766	Q62766 ratius norv
27	30	100.0	1596	11	Q921F7	Q921F7 ratius norv
28	30	100.0	2524	5	Q9CRA5	Q9CRA5
29	30	100.0	2524	5	Q9CRA5	Q9CRA5
30	30	100.0	103	2	Q9C9M3	Q9C9M3
31	30	100.0	211	10	Q9C6G9	Q9C6G9
32	30	100.0	252	10	Q9LZB5	Q9LZB5
33	30	100.0	254	10	Q9HLK2	Q9HLK2 thermoplasma
34	30	100.0	267	2	Q53815	Q53815 streptomyces
35	30	100.0	280	10	Q9PFL20	Q9PFL20 arabidopsis
36	30	100.0	288	5	Q9VBE2	Q9VBE2 drosophila
37	30	100.0	392	5	Q77475	Q77475 drosophila
38	30	100.0	421	9	Q9VD28	Q9VD28 drosophila
39	30	100.0	425	5	Q64306	Q64306 bacteriophage
40	30	100.0	425	5	Q9V7F3	Q9V7F3 drosophila
41	30	100.0	455	13	Q07426	Q07426 carassius a
42	30	100.0	464	5	Q61610	Q61610 acropora m
43	30	100.0	504	10	Q80655	Q80655 arabidopsis
44	30	100.0	562	1	Q9YCC5	Q9YCC5 aeropyrum p
45	30	100.0	569	10	Q9SL10	Q9SL10 arabidopsis
46	30	100.0	571	2	Q91502	Q91502 pseudomonas
47	30	100.0	571	2	Q9C172	Q9C172 pasteurella
48	30	100.0	593	5	Q9B829	Q9B829 ceratitidis c
49	30	100.0	597	5	Q25658	Q25658 plasmidium
50	30	100.0	619	10	Q9FMO0	Q9FMO0 arabidopsis
51	30	100.0	654	5	Q16814	Q16814 drosophila
52	30	100.0	675	10	Q9ML83	Q9ML83 arabidopsis
53	30	100.0	676	5	Q9V1X5	Q9V1X5 drosophila
54	30	100.0	729	10	Q9ZU22	Q9ZU22 arabidopsis
55	30	100.0	749	10	Q9C6N5	Q9C6N5 arabidopsis
56	30	100.0	851	2	Q91442	Q91442 pseudomonas
57	30	100.0	882	10	Q65231	Q65231 arabidopsis
58	30	100.0	893	10	Q9LHT1	Q9LHT1 arabidopsis
59	30	100.0	895	10	Q80973	Q80973 arabidopsis
60	30	100.0	940	4	Q00405	Q00405 homo sapien
61	30	100.0	941	10	Q9JTH0	Q9JTH0 arabidopsis
62	30	100.0	943	10	Q9SEF7	Q9SEF7 arabidopsis
63	30	100.0	1028	10	Q9ZPF2	Q9ZPF2 arabidopsis
64	30	100.0	1086	5	Q9VW31	Q9VW31 drosophila
65	30	100.0	1127	10	Q9C608	Q9C608 arabidopsis
66	30	100.0	1166	11	Q9ET81	Q9ET81 ratius norv
67	30	100.0	1171	11	Q9QX12	Q9QX12 ratius norv
68	30	100.0	1179	4	Q9QGE2	Q9QGE2 homo sapien
69	30	100.0	1249	11	Q9QX02	Q9QX02 ratius norv
70	30	100.0	1284	11	Q9ESK6	Q9ESK6 ratius norv
71	30	100.0	1293	11	Q9QUH6	Q9QUH6 ratius norv
72	30	100.0	1297	5	Q9V604	Q9V604 drosophila
73	30	100.0	70	6	Q9XSS1	Q9XSS1 canis famill
74	30	100.0	82	4	Q9U0R5	Q9U0R5 homo sapien
75	30	100.0	109	5	Q9VEE3	Q9VEE3 drosophila
76	30	100.0	111	12	Q9VW69	Q9VW69 prospect hi
77	30	100.0	115	10	Q9M7D7	Q9M7D7 pisum sativ
78	30	100.0	126	4	Q9H97	Q9H97 homo sapien
79	30	100.0	178	5	Q9B1S7	Q9B1S7 tetragonath
80	30	100.0	179	2	Q9F2U1	Q9F2U1 streptomyces
81	30	100.0	180	2	Q53453	Q53453 bos taurus
82	30	100.0	181	2	Q87052	Q87052 mycobacteri
83	30	100.0	181	2	Q87052	Q87052 vibrio chol
84	30	100.0	182	12	Q9HPC2	Q9HPC2 homo sapien
85	30	100.0	192	2	Q9KH83	Q9KH83 meleagrid h
86	30	100.0	192	2	Q9KH88	Q9KH88 streptomyces
87	30	100.0	199	12	Q75026	Q75026 human immun
88	30	100.0	204	12	Q75025	Q75025 human immun
89	30	100.0	205	12	Q75023	Q75023 human immun
90	30	100.0	207	12	Q75016	Q75016 human immun
91	30	100.0	208	12	Q75017	Q75017 human immun
92	30	100.0	219	5	Q9VYN7	Q9VYN7 drosophila
93	30	100.0	231	10	Q65324	Q65324 raphanus sa

93 27 90.0 240 5 Q9W396  
94 27 90.0 242 1 Q9PCP1  
95 27 90.0 245 1 Q9H015  
96 27 90.0 246 1 Q9J0M9  
97 27 90.0 251 5 Q9VM28  
98 27 90.0 251 11 P70331  
99 27 90.0 254 10 Q9LD63  
100 27 90.0 256 5 Q9V5C1  
Q9V5C1 drosophila

## ALIGNMENTS

RESULT 1  
Q9YD26 PRELIMINARY: PRT: 107 AA.  
ID Q9YD26  
AC Q9YD26:  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE HYPOTHEORETICAL 11.4 KDA PROTEIN APE0773.  
GN APE0773.  
OS Aeropyrum pernix.  
OC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;  
OC Aeropyrum.  
OX NCBI\_TaxID=56636;  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K1.  
RX MEDLINE=99310339; PubMed=10382966;  
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Halkawa Y.,  
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Anka A., Kosugi H.,  
RA Hasegawa A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,  
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,  
RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,  
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;  
RT "Complete genome sequence of an aerobic hyper-thermophilic  
RT crenarchaeon, Aeropyrum pernix K1.";  
RL DNA Res. 6:83-101(1999).  
DR EMBL: AP000060; BAA79751.1; -  
KW Hypothetical protein; Complete proteome  
SQ SEQUENCE 107 AA; 11356 MW; E26C32C0A372D831 CRC64;

Query Match 100.0%; Score 30; DB 1; Length 107;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLGGS 6  
Db 63 GSLGGS 68

RESULT 2  
Q9VBD9 PRELIMINARY: PRT: 175 AA.  
ID Q9VBD9  
AC Q9VBD9:  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
DE CG6447 PROTEIN.  
GN CG6447.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY.  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amaratilake P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.C., Mortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Belt G., Nelson C.R., Miklos G.L.G.,  
RA Bailly J.F., Agbayani A., An H.-J., Andrews-Plamkock C., Baldwin D.,  
RA Barille R.M., Basu A., Baxendale J., Bayraktaroglu I., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Borchan M.R., Bouck J., Brookslein P., Brothier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cavley S., Dahke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodet A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegami C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zavel J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zibbs R.A., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
DR EMBL: AE003757; AAF56603.1; -  
DR FlyBase: FBgn0039437; CG6447.  
SQ SEQUENCE 175 AA; 16827 MW; 96A83146DF8C8BF8 CRC64;

Query Match 100.0%; Score 30; DB 5; Length 175;  
Best Local Similarity 100.0%; Pred. No. 98;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLGGS 6  
Db 79 GSLGGS 84

RESULT 3  
Q75013 PRELIMINARY: PRT: 203 AA.  
ID Q75013  
AC Q75013:  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE ENVELOPE GLYCOPROTEIN (FRAGMENT).  
GN ENV.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TWIN A.  
RX MEDLINE=96211491; PubMed=8648693.  
RA Hutto C., Zhou Y., He J., Geffin R., Hill M., Scott W., Wood C.,  
RA "Longitudinal studies of viral sequence, viral phenotype, and  
RT immunologic parameters of human immunodeficiency virus type 1  
RT infection in perinatally infected twins with discordant disease  
RT courses.";  
RL J. Virol. 70:3589-3598(1996).  
DR EMBL: U47563; AAB06904.1; -



DR InterPro: IPR000777; GP120.  
DR Pfam: PF00516; GP120; 1.  
KW Envelope protein.  
FT NON\_TER 1  
FT NON\_TER 203  
SQ SEQUENCE 203 AA; 22521 MW; B2FED7FC90B810EB CRC64;

Query Match  
Best Local Similarity 100.0%; Score 30; DB 12; Length 203;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSIGGS 6  
DB 139 GSIGGS 144

RESULT 4  
ID 075018 PRELIMINARY; PRT; 209 AA.  
AC 075018;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DE ENVELOPE GLYCOPROTEIN (FRAGMENT).  
GN Env.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TWIN A;  
RX MEDLINE=96211491; PubMed=8648693;  
RA Hutto C., Zhou Y., He J., Giffin R., Hill M., Scott W., Wood C.;  
RT \*Longitudinal studies of viral sequence, viral phenotype, and  
RT immunologic parameters of human immunodeficiency virus type 1  
RT infection in perinatally infected twins with discordant disease  
RT courses.";  
RL J. Virol. 70:3589-3598(1996).  
DR EMBL: U47568; AAB06909.1; -;  
DR InterPro: IPR000777; GP120.  
DR Pfam: PF00516; GP120; 1.  
KW Envelope protein.  
FT NON\_TER 1  
FT NON\_TER 209  
SQ SEQUENCE 209 AA; 23290 MW; 00BD8584476EF65 CRC64;

Query Match  
Best Local Similarity 100.0%; Score 30; DB 12; Length 209;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSIGGS 6  
DB 144 GSIGGS 149

RESULT 5  
ID 061297 PRELIMINARY; PRT; 277 AA.  
AC 061297;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DE EXTRACELLULARMATRIX-LIKE PROTEIN, COMPLETE CDS.  
OS Haemaphysalis longicorais.  
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;  
OC Parasitiformes; Ixodida; Ixodidae; Haemaphysalis.  
OX NCBI\_TaxID=44386;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Mullen A., Sugimoto C., Onuma M.;  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL: AB014612; BAA28620.1; -;  
SQ SEQUENCE 277 AA; 27973 MW; 5B3297AFA3ABE09C CRC64;

Query Match  
Best Local Similarity 100.0%; Score 30; DB 5; Length 277;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSIGGS 6  
DB 39 GSIGGS 44

RESULT 6  
ID 054259 PRELIMINARY; PRT; 281 AA.  
AC 054259;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DE GLA 0 GENES STRB1, STRF, STRG, STRH, STRP.  
GN STRP.  
OS Streptomyces glaucescens.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1907;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GLA 0;  
RA Mayer G., Piepersberg W.;  
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 13032 (GLA.0);  
RA Piepersberg W.;  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 13032 (GLA.0);  
RA Retzlaff L., Mayer G., Beyer S., Ahlert J., Verseck S., Distler J.,  
RA Piepersberg W.;  
RT \*Streptomycin Production in Streptomyces: a Progress Report.";  
RL (in) Baltz R.H., Hegeman G.D., Skatrud P.L. (eds.);  
RL Industrial microorganisms. Basic and applied molecular genetics,  
RL pp.183-194, ASM Press, Herndon (1993).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 13032 (GLA.0);  
RX MEDLINE=96204519; PubMed=8628239;  
RA Beyer S., Distler J., Piepersberg W.;  
RT \*The str gene cluster for the biosynthesis of 5'-hydroxystreptomycin  
RT in Streptomyces glaucescens GLA.0 (ETH 22794): new operons and  
RT evidence for pathway-specific regulation by StrR.";  
RL Mol. Gen. Genet. 250:775-784(1996).  
DR EMBL: X78974; CAA55572.1; -;  
DR EMBL: AJ006985; CAA07379.1; -;  
SQ SEQUENCE 281 AA; 31427 MW; 11A2C5F4516EBED7 CRC64;

Query Match  
Best Local Similarity 100.0%; Score 30; DB 2; Length 281;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSIGGS 6  
DB 91 GSIGGS 96

RESULT 7  
ID 09VBE0 PRELIMINARY; PRT; 286 AA.  
AC 09VBE0;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)  
 DE CG6478 PROTEIN.  
 GN CG6478.  
 OS Drosophila melanogaster (fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_TaxID=7227;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY.  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abill J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Crawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodex A., Gong F., Correll J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibbagwam C.,  
 RA Jajall M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Meltel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Mekulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclet J.M.,  
 RA Palazzolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Sliden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskef R., Tector C., Turner K., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan X., Zhang G., Zhao Q., Zheng L.,  
 RA Zhong X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Glads R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AF003757; AAF56602.1; -  
 DR FlyBase: FBgn0039436; CG6478.  
 SO SEQUENCE 286 AA; 28617 MW; 53A4059D37D2BD4E CRC64;

Query Match 100.0%; Score 30; DB 5; Length 286;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLGGS 6  
 DB 79 GSLGGS 84

RESULT 8  
 ID 09X6F4 PRELIMINARY; PRT; 371 AA.  
 AC 09X6F4;  
 DT 01-NOV-1999 (Tremblrel. 12, Created)  
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
 DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
 DE FIRONECTIN-BINDING PROTEIN SFS.

GN SFS.  
 OS Streptococcus equi.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 NCBI\_TaxID=1336;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BD3221;  
 RX MEDLINE=99242824; PubMed=10225899;  
 RA Lindmark H., Guss B.;  
 RT "SFS, a novel fibronectin-binding protein from Streptococcus equi,  
 RT inhibits the binding between fibronectin and collagen."  
 RL Infect. Immun. 67:2383-2388(1999).  
 DR EMBL: AF136451; AAD29424.1; -  
 SO SEQUENCE 371 AA; 39765 MW; 37BA7B59A0DD8535 CRC64;

Query Match 100.0%; Score 30; DB 2; Length 371;  
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLGGS 6  
 DB 40 GSLGGS 45

RESULT 9  
 ID 09B0G1 PRELIMINARY; PRT; 590 AA.  
 AC 09B0G1;  
 DT 01-JUN-2001 (Tremblrel. 17, Created)  
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE HYPOTHEETICAL 63.3 KDA PROTEIN.  
 GN DKRP7610132.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 NCBI\_TaxID=9606;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=AMYGDALA;  
 RX MEDLINE=21154917; PubMed=11230166;  
 RA Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,  
 RA Ansoerje W., Boecker M., Bloecher H., Bauersachs S., Blum H.,  
 RA Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,  
 RA Meves H.W., Ottenwelder B., Obermaier B., Tampe J., Heubner D.,  
 RA Wandut R., Korn B., Klein M., Poustka A.;  
 RT "Towards a catalog of Human Genes and Proteins: Sequencing and  
 RT Analysis of 500 Novel Complete Protein Coding Human CDNA's."  
 RL Genome Res. 11:422-435(2001).  
 DR EMBL: AL136594; CAB66529.1; -  
 KW Hypothetical protein.  
 SO SEQUENCE 590 AA; 63303 MW; 1E6B9F2C44E7935 CRC64;

Query Match 100.0%; Score 30; DB 4; Length 590;  
 Best Local Similarity 100.0%; Pred. No. 4e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLGGS 6  
 DB 149 GSLGGS 154

RESULT 10  
 ID 09FGC2 PRELIMINARY; PRT; 750 AA.  
 AC 09FGC2;  
 DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE DNA HELICASE-LIKE.

OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 NC NCBITaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-COLDBIA;  
 RA Kaneke T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,  
 RA Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI."  
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB026643; BAB09253.1; -  
 DR InterPro: IPR001410; DEAD.  
 DR SMART: SM00487; DEXDC; 1.  
 KW Helicase.  
 SO SEQUENCE 750 AA; 81452 MW; 88B1F2D4EA8B1457 CRC64;

Query Match 100.0%; Score 30; DB 10; Length 750;  
 Best Local Similarity 100.0%; Pred. No. 5.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSLSGS 6  
 DB 733 GSLSGS 738

RESULT 11  
 ID 090193 PRELIMINARY; PRT; 763 AA.  
 AC 090193;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE HYPOTHETICAL 76.7 KDA PROTEIN.  
 GN L2903.05.  
 OS Leishmania major.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OX NCBITaxID=5664;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-FRIEDLIN;  
 RA Medler H., Duesterhoeft A., Ivens A.C., Lawson D., Quail M.,  
 RA Rajandream M.A., Barrell B.G.;  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-FRIEDLIN;  
 RX MEDLINE-98146435; PubMed-9477341;  
 RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,  
 RA Smith D.F.;  
 RT "A physical map of the Leishmania major Friedlin genome."  
 RL Genome Res. 8:135-145(1998).  
 DR EMBL: AL117319; CAB5519.1; -  
 KW Hypothetical protein.  
 SO SEQUENCE 763 AA; 78737 MW; 815A15ECA1265361 CRC64;

Query Match 100.0%; Score 30; DB 5; Length 763;  
 Best Local Similarity 100.0%; Pred. No. 5.4e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSLSGS 6  
 DB 123 GSLSGS 128

RESULT 12  
 ID 09AV46 PRELIMINARY; PRT; 773 AA.  
 AC 09AV46;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE HYPOTHETICAL 86.7 KDA PROTEIN.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzeae; Oryza.  
 OX NCBITaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. NIPPONBARE;  
 RA Buell C.R., Yuan Q., Ouyang S., Moffat K.S., Hill J.N., Burr P.C.,  
 RA Hsiao J., Zismann V., Pal G., Bowman C.L., Fujii C.Y., VanAken S.E.,  
 RA Bowman C.L., Claven B., Utterback T.R., Khalak H., Feldblyum T.V.,  
 RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;  
 RT "Oryza sativa chromosome 10 BAC OSJNB40093B11 genomic sequence."  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC024594; AAK21344.1; -  
 KW Hypothetical protein.  
 SO SEQUENCE 773 AA; 86701 MW; 8F7B56871DC91AE0 CRC64;

Query Match 100.0%; Score 30; DB 10; Length 773;  
 Best Local Similarity 100.0%; Pred. No. 5.5e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSLSGS 6  
 DB 447 GSLSGS 452

RESULT 13  
 ID 09AC84 PRELIMINARY; PRT; 802 AA.  
 AC 09AC84;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE CSUD.  
 GN CSUD.  
 OS CSUD.  
 OS Vibrio parahaemolyticus.  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
 OX NCBITaxID=670;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BB22;  
 RA McArthur L.L., Welsh C., Boles B.;  
 RT "The csu locus of Vibrio parahaemolyticus."  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF339087; AAK37523.1; -  
 SO SEQUENCE 802 AA; 88127 MW; 653711976A2BF37A CRC64;

Query Match 100.0%; Score 30; DB 2; Length 802;  
 Best Local Similarity 100.0%; Pred. No. 5.8e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSLSGS 6  
 DB 532 GSLSGS 537

RESULT 14  
 ID 044125 PRELIMINARY; PRT; 898 AA.  
 AC 044125;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE TRANSFERRIN BINDING PROTEIN 1.  
 GN TBPA.  
 OS Actinobacillus pleuropneumoniae (Haemophilus pleuropneumoniae).  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

OC Actinobacillus.  
 OX NCBI\_TaxID=715;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H49;  
 RX MEDLINE=96036198; PubMed=7582000;  
 RA Gonzalez G.C., Yu R.H., Rosteck P.R. Jr., Schryvers A.B.;  
 RT "Sequence, genetic analysis, and expression of Actinobacillus  
 pleuropneumoniae transferrin receptor genes."  
 RL Microbiology 141:2405-2416(1995).  
 DR EMBL: U16017; AAC43485.1; -;  
 DR InterPro: IPR000531; TonB\_boxC.  
 DR Pfam: PF00593; TonB\_boxC; 1.  
 DR PROSITE: PS01156; TONB\_DEPENDENT\_REC\_2; 1.  
 SO SEQUENCE 898 AA; 102187 MW; 4E49988810A3752A CRC64;

Query Match 100.0%; Score 30; DB 2; Length 898;  
 Best Local Similarity 100.0%; Pred. No. 6.6e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLGGS 6  
 |||||  
 DB 151 GSLGGS 156

RESULT 15

ID 044158 PRELIMINARY; PRT; 898 AA.  
 AC 044158;  
 DT 01-NOV-1996 (TREMBLREL. 01, Created)  
 DT 01-NOV-1996 (TREMBLREL. 01, Last sequence update)  
 DT 01-JUN-2001 (TREMBLREL. 17, Last annotation update)  
 DE TRANSFERRIN RECEPTOR.  
 GN TRFB.  
 OS Actinobacillus pleuropneumoniae (Haemophilus pleuropneumoniae).  
 OG Plasmid pPF205/023.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 CC Actinobacillus.  
 OX NCBI\_TaxID=715;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AP205;  
 RA Wilke M.;  
 RL Thesis (1995). Institut fuer Mikrobiologie,  
 Tierärztliche Hochschule Hannover, Germany.  
 DR EMBL: Z54191; CA90896.1; -;  
 DR InterPro: IPR000531; TonB\_boxC.  
 DR Pfam: PF00593; TonB\_boxC; 1.  
 DR PROSITE: PS01156; TONB\_DEPENDENT\_REC\_2; 1.  
 KW Plasmid.  
 SO SEQUENCE 898 AA; 102217 MW; AD627D58C0B7E6E2 CRC64;

Query Match 100.0%; Score 30; DB 2; Length 898;  
 Best Local Similarity 100.0%; Pred. No. 6.6e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLGGS 6  
 |||||  
 DB 151 GSLGGS 156

RESULT 16

ID 044127 PRELIMINARY; PRT; 931 AA.  
 AC 044127;  
 DT 01-NOV-1996 (TREMBLREL. 01, Created)  
 DT 01-NOV-1996 (TREMBLREL. 01, Last sequence update)  
 DT 01-JUN-2001 (TREMBLREL. 17, Last annotation update)  
 DE TRANSFERRIN BINDING PROTEIN 1.  
 GN TRBA.  
 OS Actinobacillus pleuropneumoniae (Haemophilus pleuropneumoniae).  
 SO SEQUENCE 931 AA; 106326 MW; D77366E2FDA54F9 CRC64;

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 CC Actinobacillus.  
 OX NCBI\_TaxID=715;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H171;  
 RX MEDLINE=96036198; PubMed=7582000;  
 RA Gonzalez G.C., Yu R.H., Rosteck P.R. Jr., Schryvers A.B.;  
 RT "Sequence, genetic analysis, and expression of Actinobacillus  
 pleuropneumoniae transferrin receptor genes."  
 RL Microbiology 141:2405-2416(1995).  
 DR EMBL: U16019; AAC43487.1; -;  
 DR InterPro: IPR000531; TonB\_boxC.  
 DR Pfam: PF00593; TonB\_boxC; 1.  
 DR PROSITE: PS01156; TONB\_DEPENDENT\_REC\_2; 1.  
 SO SEQUENCE 931 AA; 106369 MW; CE88D077D03C54F9 CRC64;

Query Match 100.0%; Score 30; DB 2; Length 931;  
 Best Local Similarity 100.0%; Pred. No. 6.9e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLGGS 6  
 |||||  
 DB 151 GSLGGS 156

RESULT 17

ID 044166 PRELIMINARY; PRT; 931 AA.  
 AC 044166;  
 DT 01-NOV-1996 (TREMBLREL. 01, Created)  
 DT 01-NOV-1996 (TREMBLREL. 01, Last sequence update)  
 DT 01-JUN-2001 (TREMBLREL. 17, Last annotation update)  
 DE TBP1 GENE PRECURSOR.  
 GN TBP1.  
 OS Actinobacillus pleuropneumoniae (Haemophilus pleuropneumoniae).  
 OG Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 CC Actinobacillus.  
 OX NCBI\_TaxID=715;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 27088;  
 RX MEDLINE=96207589; PubMed=8670116;  
 RA Daban M., Medrano A., Querol E.;  
 RT "Cloning, sequencing and expression of the transferrin-binding protein  
 RT 1 gene from Actinobacillus pleuropneumoniae."  
 RL Biochem. J. 315:257-264(1996).  
 DR EMBL: Z49708; CA89810.1; -;  
 DR InterPro: IPR000531; TonB\_boxC.  
 DR Pfam: PF00593; TonB\_boxC; 1.  
 DR PROSITE: PS01156; TONB\_DEPENDENT\_REC\_2; 1.  
 KW Signal.  
 FT SIGNAL 1 22  
 FT CHAIN 23 931 TBP1 GENE.  
 SO SEQUENCE 931 AA; 106326 MW; D77366E2FDA54F9 CRC64;

Query Match 100.0%; Score 30; DB 2; Length 931;  
 Best Local Similarity 100.0%; Pred. No. 6.9e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLGGS 6  
 |||||  
 DB 151 GSLGGS 156

RESULT 18

ID 09P620 PRELIMINARY; PRT; 935 AA.  
 AC 09P620;  
 DT 01-OCT-2000 (TREMBLREL. 15, Created)  
 DT 01-OCT-2000 (TREMBLREL. 15, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)  
DE CONSERVED HYPOTHETICAL PROTEIN.  
GN 13E11.260.  
OS Neurospora crassa.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariales; Sordariaceae; Neurospora.  
OX NCBI\_TaxID=5141;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Schulte U., Aign V., Hohelsel J., Brandt P., Fartmann B., Holland R.,  
RA Nyakatura G., Mewes H.W., Mannhaupt G.;  
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA German Neurospora genome project;  
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AL353820; CAB8594.1; -;  
DR InterPro: IPR000822; Znf-C2H2.  
DR Pfam: PF00096; Znf-C2H2; 2.  
DR SMART: SM00355; Znf-C2H2; 2.  
DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 2.  
DR PROSITE: PS00157; ZINC\_FINGER\_C2H2\_2; 2.  
KW DNA-binding; Metal-binding; Zinc-finger.  
SQ SEQUENCE 935 AA; 103980 MW; 35F0F6450AC2ED20 CRC64;

Query Match 100.0%; Score 30; DB 3; Length 935;  
Best Local Similarity 100.0%; Pred. No. 6.9e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSIGGS 6  
Db 700 GSIGGS 705

RESULT 19  
Q9R2S3 PRELIMINARY; PRT; 936 AA.  
AC Q9R2S3;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)  
DE HYPOTHETICAL 97.7 KDA PROTEIN.  
GN DRB0041.  
OS Deinococcus radiodurans.  
OC Plasmid Mpi.  
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.  
OX NCBI\_TaxID=1293;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-R1;  
RX MEDLINE-20036896; PubMed-10567266;  
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
RA Dodson R.V., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
RA Voita K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
RA Vamathevan J.J., Lam P., McDonald L., Uitterback T., Zaleski C.,  
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
RA Fraser C.M.;  
RT "Genome sequence of the radioresistant bacterium Deinococcus  
radiodurans R1";  
RL Science 286:1571-1577(1999).  
DR EMBL: AE001826; AAF12631.1; -;  
DR TIGR: DRB0041; -;  
DR InterPro: IPR002173; PFKB.  
DR PROSITE: PS00583; PFKB\_KINASES\_1; UNKNOWN\_1.  
KW Hypothetical protein; Plasmid; Complete proteome.  
SQ SEQUENCE 936 AA; 97723 MW; 4ABD43D7F1495C0 CRC64;

Query Match 100.0%; Score 30; DB 2; Length 936;  
Best Local Similarity 100.0%; Pred. No. 6.9e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSIGGS 6  
Db 800 GSIGGS 805

RESULT 20  
O9K1V0 PRELIMINARY; PRT; 993 AA.  
ID O9K1V0  
AC O9K1V0;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)  
DE HEMOGLOBIN BINDING PROTEIN C.  
GN HGBC.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-N182;  
RX MEDLINE-20316037; PubMed-10858226;  
RA Cope L.D., Hirkal Z., Hansen E.J.;  
RT "Detection of Phase Variation in Expression of Proteins Involved in  
RT Hemoglobin and Hemoglobin-Haptoglobin Binding by Nontypeable  
RT Haemophilus influenzae";  
RL Infect. Immun. 68:4092-4101(2000).  
DR EMBL: AF221060; AAF80178.1; -;  
DR InterPro: IPR001993; Mitoch\_carrier.  
DR InterPro: IPR000531; TonB\_boxC.  
DR Pfam: PF00593; TonB\_boxC; 1.  
DR PROSITE: PS00215; MITOCH\_CARRIER; UNKNOWN\_1.  
DR PROSITE: PS01156; TONB\_DEPENDENT\_REC\_2; 1.  
SQ SEQUENCE 993 AA; 113616 MW; A551BF3B2C641612 CRC64;

Query Match 100.0%; Score 30; DB 2; Length 993;  
Best Local Similarity 100.0%; Pred. No. 7.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSIGGS 6  
Db 174 GSIGGS 179

RESULT 21  
O87296 PRELIMINARY; PRT; 999 AA.  
ID O87296  
AC O87296;  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)  
DE HEMOGLOBIN BINDING PROTEIN.  
GN HGFB.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H1689 TYPE B;  
RX MEDLINE-98427137; PubMed-9746572;  
RA Ken Z., Jin H., Morton D.J., Stull T.L.;  
RT "HgbB, a gene encoding a second Haemophilus influenzae hemoglobin- and  
RT hemoglobin-haptoglobin-binding protein";  
RL Infect. Immun. 66:4733-4741(1998).  
DR EMBL: AF022910; AAC60790.1; -;  
DR InterPro: IPR000531; TonB\_boxC.  
DR InterPro: IPR001993; Mitoch\_carrier.  
DR Pfam: PF00593; TonB\_boxC; 1.  
DR PROSITE: PS00215; MITOCH\_CARRIER; UNKNOWN\_1.  
DR PROSITE: PS01156; TONB\_DEPENDENT\_REC\_2; 1.

SQ SEQUENCE 999 AA; 114435 MW; 58F631FA5D2685B0 CRC64;  
 Query Match 100.0%; Score 30; DB 2; Length 999;  
 Best Local Similarity 100.0%; Pred. No. 7.4e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GSLGCS 6  
 |||||  
 DB 193 GSLGCS 198  
 RESULT 22  
 Q9KIV2 PRELIMINARY; PRT; 1013 AA.  
 AC Q9KIV2:  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE HEMOGLOBIN BINDING PROTEIN A.  
 GN HGBA.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus.  
 NCBI\_TaxID=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-N182;  
 RX MEDLINE-20316037; PubMed=10856226;  
 RA Cope L.D., Hkhal Z., Hansen E.J.;  
 RT \*Detection of Phase Variation in Expression of Proteins Involved in  
 RT Hemoglobin and Hemoglobin-Haploglobin Binding by Nontypable  
 RT Haemophilus influenzae.\*;  
 RL Infect. Immun. 68:4092-4101(2000).  
 DR EMBL: AF221059; AAE80176.1; -.  
 DR InterPro: IPR001993; Mitoch\_carrier.  
 DR InterPro: IPR000531; TonB\_boxC.  
 DR Pfam: PF00593; TonB\_boxC; 1.  
 DR PROSITE: PS00215; MITOCH\_CARRIER; UNKNOWN\_1.  
 DR PROSITE: PS01156; TONB\_DEPENDENT\_REC\_2; 1.  
 DR SEQUENCE 1013 AA; 116260 MW; 769964335A4ED3C1 CRC64;

Query Match 100.0%; Score 30; DB 2; Length 1013;  
 Best Local Similarity 100.0%; Pred. No. 7.6e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GSLGCS 6  
 |||||  
 DB 194 GSLGCS 199  
 RESULT 23  
 Q9VGA4 PRELIMINARY; PRT; 1081 AA.  
 AC Q9VGA4:  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE CG10042 PROTEIN.  
 GN CG10042 OR CG17393.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY;  
 RX MEDLINE-20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 Ramanidhes P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 " 9

RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champé M., Pfeiffer B.D.,  
 RA Man K.H., Doyle C., Baxter E.G., Holt G., Nelson C.R., Miklos G.L.G.,  
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
 RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Dey A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ileguam C.,  
 RA Jatal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheefer F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Wodgett T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gbbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT \*The genome sequence of Drosophila melanogaster.\*;  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003695; AAF54781.1; -.  
 DR FlyBase: FBgn0038016; CG10042.  
 DR InterPro: IPR001739; PHD.  
 DR InterPro: IPR001965; PHD.  
 DR InterPro: IPR002998; Tudor.  
 DR InterPro: IPR008822; ZnF-C2H2.  
 DR Pfam: PF01429; MBD; 1.  
 DR Pfam: PF00628; PHD; 1.  
 DR Pfam: PF00096; zf-C2H2; 1.  
 DR SMART: SM00391; MBD; 1.  
 DR SMART: SM00249; PHD; 1.  
 DR SMART: SM00333; TUDOR; 2.  
 DR SMART: SM00355; ZnF\_C2H2; 1.  
 DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 1.  
 DR PROSITE: PS0157; ZINC\_FINGER\_C2H2\_2; 1.  
 KW DNA-binding; Metal-binding; Zinc-finger.  
 SQ SEQUENCE 1081 AA; 119852 MW; D3417492763DD508 CRC64;

Query Match 100.0%; Score 30; DB 5; Length 1081;  
 Best Local Similarity 100.0%; Pred. No. 8.2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GSLGCS 6  
 |||||  
 DB 287 GSLGCS 292  
 RESULT 24  
 Q9VHR2 PRELIMINARY; PRT; 1340 AA.  
 AC Q9VHR2:  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE CG11729 PROTEIN.  
 GN CG11729.  
 OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celinker S.E., Holt R.A., Hoskins R.A., Galle R.F.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Bratton R.C., Rogers Y.H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abrell J.F., Abayaratne A., An H.-J., Andrews-Pfankuch C., Baldwin D.,  
 RA Baller R.M., Basu A., Baxendale J., Bayraktoglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,  
 RA Borrova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
 RA Butris K.C., Busam D.A., Butler C., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwa C.,  
 RA Jalali B., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasro P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spitzer E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weissstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*."  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003678; AAF4239.2; -  
 DR FlyBase: FBgn0037586; CG11729.  
 SQ SEQUENCE 1340 AA; 147657 MW; 8E2AC5364C4CA0A0 CRC64;

Query Match 100.0%; Score 30; DB 5; Length 1340;  
 Best Local Similarity 100.0%; Pred. NO. 1e+03;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLGGS 6  
 DB 637 GSLGGS 642

RESULT 25  
 ID 060976 PRELIMINARY; PRT; 1557 AA.  
 AC 060976;  
 DT 01-AUG-1998 (TREMBlrel. 07, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE STPK1.  
 GN STPK1.  
 OS Leishmania major.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OX NCBI\_TaxID=5664;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN-FRIEDLIN;  
 RA Myler P.J., Audleman L., Hixson G., Kiser P., Lemley C., Rickel E.,  
 RA Sisk E., Sunkin S., Swartzell S., Westlake T., Magness C., Bastien P.,  
 RA Fu G., Ivens A., Stuart K.;  
 RT "Leishmania major Friedlin chromosome 1 has only two polycistronic  
 RT units of protein coding genes."  
 RL Submitted (May-1998) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-FRIEDLIN;  
 RA Myler P.J.;  
 RT "The nucleotide sequence of *Leishmania major* Friedlin chromosome 1."  
 RL Submitted (May-1998) to the EMBL/Genbank/DBJ databases.  
 CC -1 SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DR EMBL: AE001274; AAC24682.1; -  
 DR InterPro: IPR000719; Euk-kinase.  
 DR InterPro: IPR002290; Ser-thr\_kin\_actsite.  
 DR Pfam: PF00069; pkinase; 4.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.  
 KW ATP-binding; Serine/threonine-protein kinase; Transferase.  
 SQ SEQUENCE 1557 AA; 167982 MW; 8C695561B8B086A CRC64;

Query Match 100.0%; Score 30; DB 5; Length 1557;  
 Best Local Similarity 100.0%; Pred. NO. 1.2e+03;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLGGS 6  
 DB 158 GSLGGS 163

RESULT 26  
 ID 062766 PRELIMINARY; PRT; 1596 AA.  
 AC 062766;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE SSECs.  
 DT 322.  
 GN Rattus norvegicus (Rat).  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-FISHER; TISSUE-EMBRYO;  
 RX MEDLINE=95257957; PubMed=7739556;  
 RA Lin X., Nelson P.J., Frankfort B., Tomblar E., Johnson R.,  
 RA Gelman I.H.;  
 RT "Isolation and characterization of a novel mitogenic regulatory gene,  
 RT 322, which is transcriptionally suppressed in cells transformed by src  
 RT and ras."  
 RL MOL. CELL. BIOL. 15:2754-2762(1995).  
 DR EMBL: U23146; AAF9517.1; -  
 DR InterPro: IPR001573; PKin\_a\_nch.  
 SQ SEQUENCE 1596 AA; 172498 MW; CF78FD2D16D9C7E2 CRC64;

Query Match 100.0%; Score 30; DB 11; Length 1596;  
 Best Local Similarity 100.0%; Pred. NO. 1.3e+03;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLGGS 6  
 DB 1347 GSLGGS 1352

RESULT 27  
 ID 0921F7

ID Q921F7 PRELIMINARY; PRT; 1596 AA.  
AC Q921F7;  
DT 01-MAY-1999 (TREMBlrel. 10, Created)  
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE PKC BINDING PROTEIN AND SUBSTRATE.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Chapline C., Mousseau B., Ramsay K., Jaken S.;  
RT "Identification of a major protein kinase C binding protein and  
RT substrate in rat embryo fibroblasts. Decreased expression in  
RT transformed cells.";  
RL J. Biol. Chem. 0:0-0(1995).  
DR EMBL: U41453; AND03788.1; -;  
DR InterPro: IPR001573; PfamA\_anch.  
SQ SEQUENCE 1596 AA; 172411 MW; E6C4FA074F1725D3 CRC64;

Query Match 100.0%; Score 30; DB 11; Length 1596;  
Best Local Similarity 100.0%; Pred. No. 1.3e+03;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GSIGS 6  
DB 1347 GSIGS 1352

RESULT 28  
Q9GPA5 PRELIMINARY; PRT; 2524 AA.

AC Q9GPA5;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE PUTATIVE NOTCH RECEPTOR PROTEIN.  
GN NOTCH.  
OS Branchiostoma floridae (Florida lancelet) (Amphioxus).  
OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;  
OC Branchiostoma.  
OX NCBI\_TaxID=7739;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-WHOLE LARVAE;  
RA Holland L.Z., Burgdorf C., Holland N.D., Lehrach H., Tamme R.,  
RA Abi-Rached L., Pontarotti P., Lardelli M.;  
RT "Cloning and developmental expression of the amphioxus homologue of  
RT Notch (Amphlnotch): evolutionary conservation of multiple expression  
RT domains in amphioxus and vertebrates.";  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: Y12539; CAC19873.1; -;  
DR InterPro: IPR002110; ANK.  
DR InterPro: IPR00152; Asx\_hydroxyl.  
DR InterPro: IPR000561; EGF-like.  
DR InterPro: IPR000742; EGF\_2.  
DR InterPro: IPR001881; EGF\_Ca.  
DR InterPro: IPR001436; EGF\_T1.  
DR InterPro: IPR002049; Laminin\_EGF.  
DR InterPro: IPR000800; Notch.  
DR Pfam: PF00008; EGF\_36.  
DR Pfam: PF00066; notch\_3.  
DR PRINTS: PRO0010; EGFBL00D.  
DR PRINTS: PRO0011; EGF\_LAMININ.  
DR PRINTS: PRO1452; NOTCH.  
DR SMART: SM00248; ANK; 6.  
DR SMART: SM00181; EGF; 37.  
DR SMART: SM00179; EGF\_Ca; 34.  
DR SMART: SM00001; EGF\_like; 13.  
DR SMART: SM00004; NL; 3.

DR PROSITE: PS50088; ANK\_REPEAT; 4.  
DR PROSITE: PS50297; ANK\_REPEAT\_REGION; 1.  
DR PROSITE: PS00010; ASX\_HYDROXYL; 1.  
KW EGF-like domain; Glycoprotein; Hydroxylation; Receptor.  
SQ SEQUENCE 2524 AA; 270969 MW; C2CA57E306D3BEC9 CRC64;

Query Match 100.0%; Score 30; DB 5; Length 2524;  
Best Local Similarity 100.0%; Pred. No. 2.2e+03;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSIGS 6  
DB 2182 GSIGS 2187

RESULT 29

Q9A9M3 PRELIMINARY; PRT; 103 AA.

AC Q9A9M3;  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE FLAGELLAR HOOK-BASAL BODY COMPLEX PROTEIN FLIE.  
GN CC0955.  
OS Caulobacter crescentus.  
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
OC Caulobacter.  
OX NCBI\_TaxID=69394;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-21173698; PubMed-11259647;  
RA Nieman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Moddock J.R.,  
RA Potocka T., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
RA Desoy R.T., Dodson R.J., Durkin A.S., Gwin M.L., Haft D.H.,  
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,  
RA Utecherback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,  
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
RT "Complete genome sequence of Caulobacter crescentus.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
DR EMBL: AE005772; ANK22939.1; -;  
DR TIGR: CC0955; -;  
KW Complete proteome.  
SQ SEQUENCE 103 AA; 10475 MW; E1F938C5A0BD6AAA CRC64;

Query Match 93.3%; Score 28; DB 2; Length 103;  
Best Local Similarity 83.3%; Pred. No. 1.4e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSIGS 6  
DB 23 GSIGS 28

RESULT 30

Q9CG69 PRELIMINARY; PRT; 211 AA.

AC Q9CG69;  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE MUTATOR-LIKE TRANSPOSASE, PUTATIVE.  
GN T1813.1.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CV. COLUMBIA;



RX MEDLINE-21016719; PubMed-11130712;  
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,  
 RA White O., Alonso J., Altati H., Araujo R., Bowman C.L., Brooks S.Y.,  
 RA Buehler E., Chao Q., Chen H., Cheuk R.F., Chin C.W.,  
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Cressy T.H., Dewar K.,  
 RA Dunn P., Egu P., Feldblyum T.V., Feng J.-D., Feng B., Fujii C.Y.,  
 RA Gili J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hultzer L.,  
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,  
 RA Kim C.J., Koo H.L., Kremetska I., Kurtz D.B., Kwan A., Lam B.,  
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,  
 RA Lin X., Liu S.X., Liu Z.A., Lueros J.S., Maizel R., Marshall A.,  
 RA Miltner J., Miranda M., Nguyen M., Nieman W.C., Osborne B.I.,  
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,  
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,  
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,  
 RA Utterback T., Van Aken S., Venterberg M., Vysotskaia V.S., Walker M.,  
 RA Wu D., Yu G., Fraser C.M., Vayberg J.C., Davis R.W.,  
 RT \*Sequence and analysis of chromosome 1 of the plant Arabidopsis  
 thaliana.\*  
 RL Nature 408:816-820(2000).  
 DR EMBL: AC079287; AAG50848.1;  
 SQ SEQUENCE 211 AA; 23856 MW; 728C76CE9AF285AB CRC64;

Query Match 93.3%; Score 28; DB 10; Length 211;  
 Best Local Similarity 83.3%; Pred. No. 3.1e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GSLGGS 6  
 Db 150 GSGGGS 155

RESULT 31  
 Q9LZB5 PRELIMINARY; PRT; 252 AA.  
 AC Q9LZB5;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE HYPOTHETICAL 28.4 KDA PROTEIN.  
 GN F8F6.160.  
 GN Arabidopsis thaliana (Mouse-ear cress).  
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eustoidia II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bevan M., Peters S.A., van Staveren M., Dirkse W., Stekema W.,  
 RA Bancroft I., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL162873; CAB85515.1;  
 KW Hypothetical protein.  
 SQ SEQUENCE 252 AA; 28426 MW; 83C716AFB88B86E CRC64;

Query Match 93.3%; Score 28; DB 10; Length 252;  
 Best Local Similarity 83.3%; Pred. No. 3.9e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GSLGGS 6  
 Db 154 GSGGGS 159

RESULT 32  
 Q9HLK2 PRELIMINARY; PRT; 254 AA.

AC Q9HLK2;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE HYPOTHETICAL PROTEIN TA0226.  
 GN TA0226.  
 OS Thermoplasma acidophilum.  
 OC Archaea; Euryarchaeota; Thermoplasmatales; Thermoplasmataceae;  
 OC Thermoplasma.  
 OX NCBI\_TaxID=2303;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-DSM 1728;  
 RX MEDLINE-20479972; PubMed-11029001;  
 RA Ruepp A., Gral W., Santos-Martinez M.-L., Koretke K.K., Volker C.,  
 RA Mewes H.-W., Fishman D., Stocker S., Lupas A.N., Baumeister W.;  
 RT \*The genome sequence of the thermophilic scavenger Thermoplasma  
 acidophilum.\*  
 RT acdophilum.  
 RL Nature 407:508-513(2000).  
 DR EMBL: AL445063; CAC11371.1;  
 KW Hypothetical protein, complete proteome.  
 SQ SEQUENCE 254 AA; 28322 MW; 241FC11271812D31 CRC64;

Query Match 93.3%; Score 28; DB 1; Length 254;  
 Best Local Similarity 83.3%; Pred. No. 3.9e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GSLGGS 6  
 Db 73 GSGGGS 78

RESULT 33  
 Q53815 PRELIMINARY; PRT; 267 AA.  
 AC Q53815;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last annotation update)  
 DE ISF 3564 GENES STRB AND STRF (FRAGMENT).  
 GN STRF.  
 OS Streptomyces blausensis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=33897;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ISP 5564;  
 RA Mayer G., Piepersberg W.;  
 RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: X78972; CAA5568.1;  
 FT NON\_TER 267  
 SQ SEQUENCE 267 AA; 29777 MW; D0D25243D0A6C3E0 CRC64;

Query Match 93.3%; Score 28; DB 2; Length 267;  
 Best Local Similarity 83.3%; Pred. No. 4.1e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GSLGGS 6  
 Db 91 GSGGGS 96

RESULT 34  
 Q9FL20 PRELIMINARY; PRT; 280 AA.  
 ID Q9FL20;  
 AC Q9FL20;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE SIMILARITY TO MOTATOR-LIKE TRANSPOSASE.

```

OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae:
OC eucosids II: Brassicales; Brassicaceae; Arabidopsis.
CX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=98344145; PubMed=9679202;
RA Kaneko T., Kotani H., Nakamura Y., Sato S., Asamizu E., Miyajima N.,
RA Tabata S.,
RT "Structural analysis of Arabidopsis thaliana chromosome 5. V. Sequence
RT features of the regions of 1,381,565 bp covered by twenty one
RT physically assigned P1 and TAC clones";
RL DNA Res. 5:131-145(1998).
DR EMBL: AB010699; BAB10892.1;
SO SEQUENCE 280 AA; 31489 MW; 852051E61DD1372 CRC64;

Query Match 93.3%; Score 28; DB 10; Length 280;
Best Local Similarity 83.3%; Pred. No. 4.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLOGS 6
DB 154 GSLOGS 159

RESULT 35
OYBE2 PRELIMINARY; PRT; 288 AA.
ID OYBE2;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE CG5468 PROTEIN.
CM CG5468.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta:
OC Pterygota: Neoptera: Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
CX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Holt R.A., Hopkins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brotlier P.,
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Chertys J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Moberly C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

```

```

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Slier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Sytkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao O.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003757; AAF56600.1;
DR Flybase: FBgn0039434; CG5468.
SO SEQUENCE 288 AA; 28669 MW; AC28B3BD4BF1B0E6 CRC64;

Query Match 93.3%; Score 28; DB 5; Length 288;
Best Local Similarity 83.3%; Pred. No. 4.5e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLOGS 6
DB 69 GSLOGS 74

RESULT 36
OYBE2 PRELIMINARY; PRT; 392 AA.
ID OYBE2;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE ANON-60AB PROTEIN.
GN CG4585 OR ANON-60AB.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta:
OC Pterygota: Neoptera: Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
CX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hopkins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Chertys J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Moberly C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Slier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

```

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Massarman D.A., Weisstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Morley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT The genome sequence of *Drosophila melanogaster*.  
 RL Science 287:2185-2195(2000).  
 RN  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99208659; PubMed=10191082;  
 RA Lukatsovich T., Asztalos Z., Juni N., Awano W., Yamamoto D.;  
 RT The *Drosophila melanogaster* 60A chromosomal division is extremely  
 RT dense with functional genes: their sequences, genomic organization,  
 RT and expression."  
 RL Genomics 57:43-56(1999).  
 DR EMBL: AF003462; AAF47081.1; -;  
 DR EMBL: AB010264; BAA32692.1; -;  
 DR EMBL: AB010261; BAA32689.1; -;  
 DR Flybase: FBgn0025335; CG4585.  
 DR InterPro: IPR000462; CDP-OH\_P\_transf.  
 KW Hypothetical protein.  
 SQ SEQUENCE 392 AA; 43857 MW; FCB6E1C454AF6140 CRC64;

Query Match 93.3%; Score 28; DB 5; Length 392;  
 Best Local Similarity 83.3%; Pred. No. 6.5e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSUGS 6  
 11:1111  
 DB 64 GSUGS 69

RESULT 37  
 O9VD28 PRELIMINARY; PRT; 408 AA.  
 AC O9VD28;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)  
 DE CG5383 PROTEIN.  
 GN CG5383.  
 OS *Drosophila melanogaster* (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC NCB1\_TaxID=7227;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY.  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celuhner S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer V., Nelson C.R., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Heit G., Champe M., Pfeiffer B.D.,  
 RA Abill J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Bailev R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Bokkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 RA Butts J.M., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Dou P.L., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,  
 RA Flocker C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Gioder A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Hoack J.,  
 RA Hostin D., Houston K.A., Howard T.J., Wei M.-H., Ibegwam C.,  
 RA Jaisl M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Murzyn D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Massarman D.A., Weisstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Morley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT The genome sequence of *Drosophila melanogaster*.  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AF003738; AAF5975.1; -;  
 DR Flybase: FBgn0038948; CG5383.  
 SQ SEQUENCE 408 AA; 47005 MW; 766B2ADC431D967B CRC64;

Query Match 93.3%; Score 28; DB 5; Length 408;  
 Best Local Similarity 83.3%; Pred. No. 6.8e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSUGS 6  
 11:1111  
 DB 400 GSUGS 405

RESULT 38  
 O64306 PRELIMINARY; PRT; 421 AA.  
 AC O64306;  
 DT 01-AUG-1998 (TREMBlrel. 07, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
 DE A-PROTEIN  
 OS bacteriophage  $\lambda$ ML1.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Leviviridae;  
 OC serogroup III.  
 GN NCB1\_TaxID=75723;  
 RN  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95239761; PubMed=7723040;  
 RA Beekwilder M.J., Nieuwenhuizen R., van Duin J.;  
 RT "Secondary structure model for the last two domains of single-stranded  
 RT RNA phage Q beta."  
 RL J. Mol. Biol. 247:903-917(1995).  
 RN  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96190948; PubMed=8609616;  
 RA Beekwilder M.J., Nieuwenhuizen R., Poot R., van Duin J.;  
 RT "Secondary structure model for the first three domains of Q beta RNA.  
 RT Control of A-protein synthesis."  
 RL J. Mol. Biol. 256:8-19(1996).  
 RN  
 RP SEQUENCE FROM N.A.  
 RA Beekwilder M.J., Nieuwenhuizen R., van Duin J.;  
 RL Submitted (Apr-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF059242; AAC14698.1; -;  
 SQ SEQUENCE 421 AA; 48486 MW; 3978CB9E7E9D0943 CRC64;

Query Match 93.3%; Score 28; DB 9; Length 421;  
 Best Local Similarity 83.3%; Pred. No. 7e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 GSUGS 6  
 11:1111  
 DB 43 GSUGS 48

RESULT 39  
 09V7F3 PRELIMINARY: PRT: 425 AA.  
 ID 09V7F3  
 AC 09V7F3  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
 DE POX-N PROTEIN.  
 GN POX-N OR CG8246.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 CX NCBI\_Taxid=7227;  
 [1]  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RA MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 Brandon R.C., Rogers Y.-H.C., Blazer G., Champe M., Pfeiffer B.D.,  
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 Beeson K.Y., Bens P.V., Bernan B.P., Bhandari D., Bolshakov S.,  
 Borokova D., Botchan M.R., Bouck J., Brockstein P., Brotler P.,  
 Butlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 Hostett D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 Jalili M., Kalush F., Karpen G.H., Ke Z., Kemison J.A., Ketchum K.A.,  
 Kimmel B.E., Kodra C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 Laake P., Lai Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclob J.M.,  
 Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 Wang Z.-Y., Wassarman D.A., Weinslock G.M., Weissenbach J.,  
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Clides R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195(2000).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 CC -1- SIMILARITY: CONTAINS A PAIRED BOX DOMAIN.  
 DR EMBL: AE003809; AAF58104.1; -  
 DR HSSP: P06601; 1PDN.  
 DR Flybase: FBgn000130; Pox-n.  
 DR InterPro: IPR001523; Paired\_box.  
 DR Pfam: PF00292; PAX; 1.  
 DR PRINTS: PR00027; PAIREDBOX.  
 DR SMART: SM00351; PAX; 1.  
 DR PROSITE: PS00034; PAIRED\_BOX; 1.  
 KW DNA-binding; Developmental protein; Nuclear protein; Paired box;  
 KW Transcription regulation.  
 SQ SEQUENCE 425 AA; 44328 MW; 59B01A4F97C1B126 CRC64;

Query Match 93.3%; Score 28; DB 5; Length 425;

Best Local Similarity 83.3%; Pred. No. 7.1e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Y 1 GSIGS 6  
 11:111  
 Db 70 GSIGS 75  
 RESULT 40  
 007426 PRELIMINARY: PRT: 455 AA.  
 ID 007426  
 AC 007426  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
 DE KERATIN.  
 OS Carassius auratus (Goldfish).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
 OC Cypriniformes; Cyprinidae; Cyprininae; Carassius.  
 CX NCBI\_Taxid=7957;  
 [1]  
 RN  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=OPTIC NERVE;  
 RA MEDLINE=94259853; PubMed=7515399;  
 RA Druger R.K., Fuchs C., Levine E.M., Park C.Y., Matthews J.P.,  
 RA Schechter N.;  
 RT "Complex expression of keratins in goldfish optic nerve."  
 RL J. Comp. Neurol. 340:269-280(1994).  
 DR EMBL: L09743; AAC38008.1; -  
 DR InterPro: IPR001664; IF.  
 DR InterPro: IPR002957; Keratin\_I.  
 DR Pfam: PF00038; filament; 1.  
 DR PRINTS: PR01248; TYPEKERATIN.  
 SQ SEQUENCE 455 AA; 49755 MW; 8C1378F2F3848EA4 CRC64;

Query Match 93.3%; Score 28; DB 13; Length 455;  
 Best Local Similarity 83.3%; Pred. No. 7.7e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Y 1 GSIGS 6  
 11:111  
 Db 14 GSIGS 19

Search completed: February 4, 2002, 08:02:24  
 Job time: 128 sec

000000

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 4, 2002, 08:01:36 ; Search time 19.95 Seconds  
(without alignments)  
6,766 Million cell updates/sec

Title: US-09-642-660-11  
Perfect score: 30  
Sequence: 1 GSI/GS 6

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues  
Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

- Database :
- 1: /cgn2\_6/prodata/2/1aa/5a\_COMB.pep.\*
  - 2: /cgn2\_6/prodata/2/1aa/5b\_COMB.pep.\*
  - 3: /cgn2\_6/prodata/2/1aa/6a\_COMB.pep.\*
  - 4: /cgn2\_6/prodata/2/1aa/6b\_COMB.pep.\*
  - 5: /cgn2\_6/prodata/2/1aa/PCTUS\_COMB.pep.\*
  - 6: /cgn2\_6/prodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	30	100.0	6	3	US-08-828-712-11
2	30	100.0	6	4	US-09-063-276-11
3	30	100.0	1346	1	US-08-635-121-2
4	28	93.3	21	1	US-08-014-1530-45
5	28	93.3	633	3	US-09-041-991A-8
6	28	93.3	633	3	US-09-041-991A-10
7	27	90.0	163	2	US-08-704-931-6
8	27	90.0	246	2	US-08-704-931-2
9	27	90.0	251	2	US-08-704-931-4
10	27	90.0	254	4	US-09-237-543-4
11	27	90.0	318	2	US-08-633-148-4
12	27	90.0	326	4	US-09-237-543-2
13	27	90.0	340	2	US-08-633-148-2
14	27	90.0	349	3	US-08-630-172-6
15	27	90.0	349	4	US-09-375-419-6
16	27	90.0	513	1	US-08-464-266-2
17	27	90.0	513	1	US-08-464-266-2
18	27	90.0	513	4	US-08-464-272-2
19	27	90.0	513	4	US-08-464-514-2
20	27	90.0	624	2	US-08-486-403-2
21	27	90.0	624	3	US-08-486-403-2
22	27	90.0	773	3	US-08-434-000A-2
23	27	90.0	773	4	US-09-312-157-2
24	27	90.0	985	2	US-08-680-326-41
25	27	90.0	985	2	US-08-687-941-13
26	27	90.0	985	2	US-08-687-941-17
27	27	90.0	985	4	US-09-074-658-13
28	27	90.0	985	4	US-09-074-658-17

## ALIGNMENTS

RESULT 1  
US-08-828-712-11  
Sequence 11, Application US/08828712  
Patent No. 6101584  
GENERAL INFORMATION:  
APPLICANT: Schneck, Jonathan P.  
APPLICANT: O'Herrin, Sean  
TITLE OF INVENTION: Soluble Divalent and Multivalent  
TITLE OF INVENTION: Heterodimeric Analogs of Proteins  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner & Witcoff, Ltd.  
STREET: 1001 G Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20001  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/828.712  
FILING DATE: 28-MAR-1997  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 01107.73713  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)508-9100  
TELEFAX: (202)508-9299  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6 amino acids  
TYPE: amino acid  
STRANDEDNESS: linear  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
FRAGMENT TYPE: Internal  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: Ig kappa peptide linker  
US-08-828-712-11

Query Match 100.0%; Score 30; DB 3; Length 6;  
Best Local Similarity 100.0%; Pred. No. 1.6e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSLGGS 6  
DB 1 GSLGGS 6

RESULT 2  
US-09-063-276-11  
Sequence 11, Application US/09063276  
Patent No. 6140113  
GENERAL INFORMATION:  
APPLICANT: Schneck, Jonathan  
APPLICANT: O'Herrin, Sean  
TITLE OF INVENTION: Molecular Complexes Which  
TITLE OF INVENTION: Modify Immune Responses  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner & Witcoff

STREET: 1001 G Street, NW  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20001  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/063.276  
FILING DATE: 21-APR-1998  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/828.712  
FILING DATE: 28-MAR-1997  
APPLICATION NUMBER: 60/014,367  
FILING DATE: 28-MAR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A  
REGISTRATION NUMBER: 32141  
REFERENCE/DOCKET NUMBER: 01107.74154  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-063-276-11

Query Match 100.0%; Score 30; DB 4; Length 6;  
Best Local Similarity 100.0%; Pred. No. 1.6e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSLGGS 6  
DB 1 GSLGGS 6

RESULT 3  
US-08-635-121-2  
Sequence 2, Application US/08635121  
Patent No. 5910442  
GENERAL INFORMATION:  
APPLICANT: Gelman, Irwin H.  
TITLE OF INVENTION: TUMOR SUPPRESSOR GENE  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond  
STREET: 30 Rockefeller Plaza  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10112-0228  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/635.121  
FILING DATE: 19-APRIL-1996  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:



NAME: Clark, Richard S  
REGISTRATION NUMBER: 26,154  
REFERENCE/DOCKET NUMBER: A30558 - 165/33603  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-408-2558  
TELEFAX: 212-765-2519  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1346 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: Internal  
ORIGINAL SOURCE:  
US-08-635-121-2

Query Match 100.0%; Score 30; DB 2; Length 1346;  
Best Local Similarity 100.0%; Pred. No. 5.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GSIGS 6  
Db 960 GSIGS 965

RESULT 4  
US-08-014-153D-45  
GENERAL INFORMATION:  
APPLICANT: Hadlock, Kenneth G.  
Foung, Steven K.H.  
Goh, Chin-Joo  
TITLE OF INVENTION: Method and Assay for HTLV  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Genelabs Technologies, Inc.  
STREET: 505 Penobscot Drive  
CITY: Redwood City  
STATE: CA  
COUNTRY: USA  
ZIP: 94063  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/014,153D  
FILING DATE: 05-Feb-1993  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/653,091  
FILING DATE: 08-FEB-1991  
APPLICATION NUMBER: US 07/366,313  
FILING DATE: 13-JUN-1989  
APPLICATION NUMBER: US 06/948,270  
FILING DATE: 31-DEC-1986  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, Allan A.  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: G4C1P3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 369-9500  
TELEFAX: (415) 368-0709  
INFORMATION FOR SEQ ID NO: 45:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 45:  
US-08-014-153D-45

Query Match 93.3%; Score 28; DB 1; Length 21;  
Best Local Similarity 83.3%; Pred. No. 16;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GSIGS 6  
Db 10 GSMGS 15

RESULT 5  
US-09-041-991A-8  
Sequence 8, Application US/09041991A  
Patent No. 6107278  
GENERAL INFORMATION:  
APPLICANT: Schnepf, H. Ernest  
APPLICANT: Narva, Kenneth E.  
APPLICANT: Muller-Cohn, Judy  
TITLE OF INVENTION: Toxins Active Against Pests  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik  
STREET: 2421 N.W. 41st Street, suite A-1  
CITY: Gainesville  
STATE: Florida  
COUNTRY: USA  
ZIP: 32606

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/041,991A  
FILING DATE: 13-MAR-1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Sanders, Jay M.  
REGISTRATION NUMBER: 39,355  
REFERENCE/DOCKET NUMBER: MA-709  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (352) 375-8100  
TELEFAX: (352) 372-5800  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 633 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-041-991A-8

Query Match 93.3%; Score 28; DB 3; Length 633;  
Best Local Similarity 83.3%; Pred. No. 5.7e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GSIGS 6  
Db 349 GSIGS 354

RESULT 6  
US-09-041-991A-10  
Sequence 10, Application US/09041991A  
Patent No. 6107278  
GENERAL INFORMATION:  
APPLICANT: Schnepf, H. Ernest

APPLICANT: Narva, Kenneth E.  
APPLICANT: Muller-Cohn, Judy  
TITLE OF INVENTION: Toxins Active Against Pests  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: Florida  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentlin  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/041,991A  
FILING DATE: 13-MAR-1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Sanders, Jay M.  
REGISTRATION NUMBER: 39,355  
REFERENCE/DOCKET NUMBER: MA-709  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (352) 375-8100  
TELEFAX: (352) 372-5800  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 633 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-041-991A-10

Query Match 93.3%; Score 28; DB 3; Length 633;  
Best Local Similarity 83.3%; Pred. No. 5.7e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSIGGS 6  
||:||||  
DB 349 GSIGGS 354

RESULT 7  
US-08-704-931-6  
Sequence 6, Application US/08704931  
Patent No. 5885797  
GENERAL INFORMATION:  
APPLICANT: Chen, Chao-Min (Amy)  
APPLICANT: Kraut, No. 5885797bert  
APPLICANT: Grouding, Mark  
APPLICANT: Weintraub, Harold  
TITLE OF INVENTION: No. 5885797el DNA Sequences Encoding Proteins  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Stratton Balliew, PLLC  
STREET: 1218 Third Avenue, Suite 1313  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentlin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/704,931  
FILING DATE:

CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: King, Jeffrey J  
REGISTRATION NUMBER: 38,515  
REFERENCE/DOCKET NUMBER: H011.P02  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-683-1496  
TELEFAX: 206-682-0446  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 163 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-704-931-6

Query Match 90.0%; Score 27; DB 2; Length 163;  
Best Local Similarity 83.3%; Pred. No. 2.1e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSIGGS 6  
||:||||  
DB 120 GALGGS 125

RESULT 8  
US-08-704-931-2  
Sequence 2, Application US/08704931  
Patent No. 5885797  
GENERAL INFORMATION:  
APPLICANT: Chen, Chao-Min (Amy)  
APPLICANT: Kraut, No. 5885797bert  
APPLICANT: Grouding, Mark  
APPLICANT: Weintraub, Harold  
TITLE OF INVENTION: No. 5885797el DNA Sequences Encoding Proteins  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Stratton Balliew, PLLC  
STREET: 1218 Third Avenue, Suite 1313  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentlin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/704,931  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: King, Jeffrey J  
REGISTRATION NUMBER: 38,515  
REFERENCE/DOCKET NUMBER: H011.P02  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-683-1496  
TELEFAX: 206-682-0446  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 246 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-704-931-2

Query Match 90.0%; Score 27; DB 2; Length 246;  
Best Local Similarity 83.3%; Pred. No. 3.3e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSLGGS 6  
Db 120 GALGGS 125

## RESULT 9

US-08-704-931-4  
; Sequence 4, Application US/08704931  
; Patent No. 5885797  
; GENERAL INFORMATION:  
; APPLICANT: Chen, Chao-Min (Amy)  
; APPLICANT: Kraut, No. 5885797bert  
; APPLICANT: Groudine, Mark  
; APPLICANT: Weintraub, Harold  
; TITLE OF INVENTION: No. 5885797el DNA Sequences Encoding Proteins  
; TITLE OF INVENTION: Involved in Myogenesis  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Stratton Ballaw, PLLC  
; STREET: 1218 Third Avenue, Suite 1313  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98101  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentln Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/704,931  
; FILING DATE:  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: King, Jeffrey J  
; REGISTRATION NUMBER: 38,515  
; REFERENCE/DOCKET NUMBER: H011.F02  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 206-683-1496  
; TELEFAX: 206-682-0446  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 251 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: Protein  
US-08-704-931-4

Query Match 90.0%; Score 27; DB 2; Length 251;  
Best Local Similarity 83.3%; Pred. No. 3.4e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GSLGGS 6  
Db 120 GALGGS 125

RESULT 10  
US-09-237-543-4  
; Sequence 4, Application US/09237543A  
; Patent No. 6143540  
; GENERAL INFORMATION:  
; APPLICANT: Kapeller, Rosana  
; TITLE OF INVENTION: NOVEL MOLECULES OF THE HKID-1-RELATED PROTEIN FAMILY  
; FILE REFERENCE: 035800/175631  
; CURRENT APPLICATION NUMBER: US/09/237,543A  
; CURRENT FILING DATE: 1999-01-26  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: Patentln Ver. 2.0  
; SEQ ID NO 4

LENGTH: 254  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: eukaryotic  
US-09-237-543-4

Query Match 90.0%; Score 27; DB 4; Length 254;  
Best Local Similarity 83.3%; Pred. No. 3.4e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSLGGS 6  
Db 41 GSLGGA 46

## RESULT 11

US-08-633-148-4  
; Sequence 4, Application US/08633148  
; Patent No. 5864018  
; GENERAL INFORMATION:  
; APPLICANT: MORSE, MICHAEL J.  
; APPLICANT: NAGASHIMA, MARIKO  
; APPLICANT: HOLLANDER, DORIS A.  
; TITLE OF INVENTION: ANTIBODIES TO ADVANCED GLYCOSYLATION  
; TITLE OF INVENTION: END-PRODUCT RECEPTOR POLYPEPTIDES AND USES THEREFOR  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: TOWNSEND & TOWNSENT & CREW LLP  
; STREET: TWO EMBARCADERO CENTER, 8TH FLOOR  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: U.S.A.  
; ZIP: 94111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentln Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/633,148  
; FILING DATE: 16-APR-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURPHY ESQ., MATTHEW B.  
; REGISTRATION NUMBER: 39,787  
; REFERENCE/DOCKET NUMBER: 014618-005600US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 326-2400  
; TELEFAX: (415) 326-2422  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 318 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-633-148-4

Query Match 90.0%; Score 27; DB 2; Length 318;  
Best Local Similarity 83.3%; Pred. No. 4.3e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSLGGS 6  
Db 309 GSVGGS 314

RESULT 12  
US-09-237-543-2

Sequence 2, Application US/09237543A  
Patent No. 6143540  
GENERAL INFORMATION:  
APPLICANT: KAPILLER, ROSANA  
TITLE OF INVENTION: NOVEL MOLECULES OF THE HKID-1-RELATED PROTEIN FAMILY  
FILE REFERENCE: 035800/175631  
CURRENT APPLICATION NUMBER: US/09/237,543A  
CURRENT FILING DATE: 1999-01-26  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 2  
LENGTH: 326  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-237-543-2

Query Match 90.0%; Score 27; DB 4; Length 326;  
Best Local Similarity 83.3%; Pred. No. 4.4e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLGCS 6  
|||||  
DB 80 GSLGCA 85

RESULT 13  
US-08-633-148-2  
Sequence 2, Application US/08633148  
Patent No. 5864018  
GENERAL INFORMATION:  
APPLICANT: MORSE, MICHAEL J.  
APPLICANT: NAGASHIMA, MARIKO  
APPLICANT: HOLLANDER, DORIS A.  
TITLE OF INVENTION: ANTIBODIES TO ADVANCED GLYCOSYLATION  
TITLE OF INVENTION: END-PRODUCT RECEPTOR POLYPEPTIDES AND USES THEREFOR  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TOWNSEND & TOWNSEND & CREW LLP  
STREET: TWO EMBARCADERO CENTER, 8TH FLOOR  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: U.S.A.  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/633,148  
FILING DATE: 16-APR-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: MURPHY ESQ., MATTHEW B.  
REGISTRATION NUMBER: 39,787  
REFERENCE/DOCKET NUMBER: 014618-0056000US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 326-2400  
TELEFAX: (415) 326-2422  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 340 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-633-148-2

Query Match 90.0%; Score 27; DB 2; Length 340;  
Best Local Similarity 83.3%; Pred. No. 4.6e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
OY 1 GSLGCS 6  
|||||  
DB 331 GSVGCS 336

RESULT 14  
US-08-630-172-6  
Sequence 6, Application US/08630172  
Patent No. 6060054  
GENERAL INFORMATION:  
APPLICANT: Staerz, Uwe  
TITLE OF INVENTION: NOVEL PRODUCT AND PROCESS FOR T  
TITLE OF INVENTION: LYMPHOCYTE VETO  
NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, 35th Floor  
CITY: Denver  
STATE: Colorado  
COUNTRY: U.S.  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/630,172  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2879-36  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 349 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-630-172-6

Query Match 90.0%; Score 27; DB 3; Length 349;  
Best Local Similarity 83.3%; Pred. No. 4.7e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GSLGCS 6  
|||||  
DB 156 GSLGCT 161

RESULT 15  
US-09-375-419-6  
Sequence 6, Application US/09375419  
Patent No. 6264950  
GENERAL INFORMATION:  
APPLICANT: Staerz, Uwe  
TITLE OF INVENTION: NOVEL PRODUCT AND PROCESS FOR T  
TITLE OF INVENTION: LYMPHOCYTE VETO  
NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, 35th Floor  
CITY: Denver  
STATE: Colorado  
COUNTRY: U.S.  
ZIP: 80203

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/375,419  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/630,172  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2879-36  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 349 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-375-419-6

Query Match 90.0%; Score 27; DB 4; Length 349;  
Best Local Similarity 83.3%; Pred. No. 4.7e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSLGGS 6  
|||||  
DB 156 GSVGCT 161

RESULT 16  
US-08-464-266-2  
Sequence 2, Application US/08464266  
Patent No. 5641652  
GENERAL INFORMATION:  
APPLICANT: ORO, Ph.D., ANTHONY E.  
APPLICANT: EVANS, Ph.D., RONALD M.  
TITLE OF INVENTION: INSECT RETINOID-LIKE RECEPTOR  
TITLE OF INVENTION: COMPOSITIONS AND METHODS  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark  
STREET: 444 South Flower Street, Suite 2000  
CITY: Los Angeles  
STATE: CA  
COUNTRY: United States  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/464,266  
FILING DATE: 05-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/013,975  
FILING DATE: 04-FEB-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/497,935  
FILING DATE: 22-FEB-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Reiter, Stephen E.  
REGISTRATION NUMBER: 31,192  
REFERENCE/DOCKET NUMBER: P41 9966

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-546-4737  
TELEFAX: 619-546-9392  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 513 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-464-266-2

Query Match 90.0%; Score 27; DB 1; Length 513;  
Best Local Similarity 83.3%; Pred. No. 7.1e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSVGGS 6  
|||||  
DB 200 GSVGGS 205

RESULT 17  
US-08-464-272-2  
Sequence 2, Application US/08464272  
Patent No. 568691  
GENERAL INFORMATION:  
APPLICANT: ORO, Ph.D., ANTHONY E.  
APPLICANT: EVANS, Ph.D., RONALD M.  
TITLE OF INVENTION: INSECT RETINOID-LIKE RECEPTOR  
TITLE OF INVENTION: COMPOSITIONS AND METHODS  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark  
STREET: 444 South Flower Street, Suite 2000  
CITY: Los Angeles  
STATE: CA  
COUNTRY: United States  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/464,272  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/013,975  
FILING DATE: 04-FEB-1993  
APPLICATION NUMBER: US 07/497,935  
FILING DATE: 22-FEB-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Reiter, Stephen E.  
REGISTRATION NUMBER: 31,192  
REFERENCE/DOCKET NUMBER: P41 9350  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-546-4737  
TELEFAX: 619-546-9392  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 513 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-464-272-2

Query Match 90.0%; Score 27; DB 1; Length 513;  
Best Local Similarity 83.3%; Pred. No. 7.1e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSVGGS 6

Db 200 GSVGCS 205

RESULT 18  
US-08-464-514-2  
Sequence 2, Application US/08464514  
Patent No. 6265173

GENERAL INFORMATION:

APPLICANT: EVANS, RONALD M.  
APPLICANT: MCKEOWN, MICHAEL B.

APPLICANT: ORO, ANTHONY E.  
APPLICANT: SEGRAVES, WILLIAM A.

APPLICANT: YAO, TSO-PANG  
TITLE OF INVENTION: MULTIMERIC FORMS OF MEMBERS OF THE  
STERIOD/THYROID SUPERFAMILY OF RECEPTORS WITH THE

TITLE OF INVENTION: ULTRASPICACLE RECEPTOR  
NUMBER OF SEQUENCES: 29

CORRESPONDENCE ADDRESS:

ADDRESSEE: PRETTY, SCHROEDER, BRUEGGEMANN & CLARK  
STREET: 444 South Flower Street, Suite 2000  
CITY: Los Angeles

STATE: California  
COUNTRY: United States

ZIP: 90071

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/464,514  
FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/907,908  
FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Reiter, Stephen E.  
REGISTRATION NUMBER: 31192

REFERENCE/DOCKET NUMBER: P41 9321  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 546-4737  
TELEFAX: (619) 546-9392

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:

LENGTH: 513 amino acids  
TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE: protein

US-08-464-514-2

Query Match

Best Local Similarity 83.3%; Pred. No. 7.1e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

1 GSLGCS 6  
DB 200 GSVGCS 205

RESULT 19

US-08-486-403-2  
Sequence 2, Application US/08486403  
Patent No. 6281330

GENERAL INFORMATION:

APPLICANT: EVANS, RONALD M.  
APPLICANT: MCKEOWN, MICHAEL B.

APPLICANT: ORO, ANTHONY E.  
APPLICANT: SEGRAVES, WILLIAM A.

APPLICANT: YAO, TSO-PANG  
TITLE OF INVENTION: MULTIMERIC FORMS OF MEMBERS OF THE

TITLE OF INVENTION: STEROID/THYROID SUPERFAMILY OF RECEPTORS WITH THE  
TITLE OF INVENTION: ULTRASPICACLE RECEPTOR  
NUMBER OF SEQUENCES: 29

CORRESPONDENCE ADDRESS:

ADDRESSEE: PRETTY, SCHROEDER, BRUEGGEMANN & CLARK  
STREET: 444 South Flower Street, Suite 2000  
CITY: Los Angeles

STATE: California  
COUNTRY: United States

ZIP: 90071

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/486,403  
FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/907,908  
FILING DATE: 02-JUL-1992

ATTORNEY/AGENT INFORMATION:

NAME: Reiter, Stephen E.  
REGISTRATION NUMBER: 31192

REFERENCE/DOCKET NUMBER: P41 9321  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 546-4737  
TELEFAX: (619) 546-9392

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:

LENGTH: 513 amino acids  
TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE: protein

US-08-486-403-2

Query Match

Best Local Similarity 83.3%; Pred. No. 7.1e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

1 GSVGCS 6  
DB 200 GSVGCS 205

RESULT 20

US-08-642-406A-22  
Sequence 22, Application US/08642406A  
Patent No. 5959177

GENERAL INFORMATION:

APPLICANT: Hehn, Mich B.  
APPLICANT: Hiatt, Andrew C.

APPLICANT: Ma, Julian K C.  
TITLE OF INVENTION: TRANSGENIC PLANTS EXPRESSING ASSEMBLED  
NUMBER OF SEQUENCES: 26

TITLE OF INVENTION: SECRETORY ANTIBODIES  
CORRESPONDENCE ADDRESS:

ADDRESSEE: THE SCRIPPS RESEARCH INSTITUTE  
STREET: 10666 No. 5959177th Torrey Pines Road, TPC-8  
CITY: La Jolla

STATE: California  
COUNTRY: US

ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/642,406A  
FILING DATE: 03-MAY-1996

CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/591,823  
FILING DATE: 02-OCT-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/427,765  
FILING DATE: 27-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Logan, April C.  
REGISTRATION NUMBER: 33,950  
REFERENCE/DOCKET NUMBER: 184.2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 554-2937  
TELEFAX: (619) 554-6312  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 624 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-642-406A-22

Query Match 90.0%; Score 27; DB 2; Length 624;  
Best Local Similarity 83.3%; Pred. No. 8.7e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSUGS 6  
DB 250 GNUGS 255

RESULT 21  
US-08-434-000A-2  
Sequence 2, Application US/08434000A  
Patent No. 6046037  
GENERAL INFORMATION:  
APPLICANT: ANDREW C. HIATT, JULIAN  
APPLICANT: K.-C. MA, THOMAS LEHNER  
TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION  
TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/434,000A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
PRIOR APPLICATION DATA: including application  
PRIOR APPLICATION DATA: described below:  
APPLICATION NUMBER: 08/367,395  
FILING DATE: 12/30/94  
ATTORNEY/AGENT INFORMATION:  
NAME: Guise, Jeffrey W.  
REGISTRATION NUMBER: 34,613  
REFERENCE/DOCKET NUMBER: 212/127  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 552-8400  
TELEFAX: (619) 552-0159  
TELEX: 67-3510

TELEX: SEQUENCE LISTING  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 773 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-434-000A-2 Rabbit polyimmunoglobulin receptor

Query Match 90.0%; Score 27; DB 3; Length 773;  
Best Local Similarity 83.3%; Pred. No. 1.1e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSUGS 6  
DB 250 GNUGS 255

RESULT 22  
US-09-312-157-2  
Sequence 2, Application US/09312157  
Patent No. 6303341  
GENERAL INFORMATION:  
APPLICANT: ANDREW C. HIATT, JULIAN  
APPLICANT: K.-C. MA, THOMAS LEHNER  
TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION  
TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/312,157  
FILING DATE: 14-May-1999  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/434,000  
FILING DATE: <unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Guise, Jeffrey W.  
REGISTRATION NUMBER: 34,613  
REFERENCE/DOCKET NUMBER: 212/127  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 552-8400  
TELEFAX: (619) 552-0159  
TELEX: 67-3510  
SEQUENCE LISTING  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 773 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
DESCRIPTION: Rabbit polyimmunoglobulin receptor  
US-09-312-157-2

Query Match 90.0%; Score 27; DB 4; Length 773;  
Best Local Similarity 83.3%; Pred. No. 1.1e+03;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSLSGS-6  
1:||||

Db 250 GSLSGS 255

RESULT 23

US-08-680-326-41

; Sequence 41, Application US/08680326  
; Patent No. 5925733

; GENERAL INFORMATION:

; APPLICANT: ROSE, TIMOTHY M.

; APPLICANT: BOSCH, MARIIX

; APPLICANT: STRAND, KURT

; APPLICANT: TODARO, GEORGE J.

; TITLE OF INVENTION: DNA POLYMERASE OF GAMMA HERPES VIRUSES

; TITLE OF INVENTION: ASSOCIATED WITH KAPOSI'S SARCOMA AND RETROPERITONEAL

; NUMBER OF SEQUENCES: 152

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORRISON & FOERSTER

; STREET: 755 Page Mill Road

; CITY: Palo Alto

; STATE: California

; COUNTRY: USA

; ZIP: 94304-1018

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/680,326

; FILING DATE:

; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:

; NAME: Schliff, J. Michael

; REGISTRATION NUMBER: 40,253

; REFERENCE/DOCKET NUMBER: 29938-20001.00

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 813-5600

; TELEFAX: (415) 494-0792

; TELEFAX: 706141

; INFORMATION FOR SEQ ID NO: 41:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 985 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-680-326-41

Query Match 90.0%; Score 27; DB 2; Length 985;

Best Local Similarity 83.3%; Pred. No. 1.4e+03;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSLSGS 6  
1:||||

Db 133 GSLSGN 138

RESULT 24

US-08-867-941-13

; Sequence 13, Application US/08867941  
; Patent No. 5977337

; GENERAL INFORMATION:

; APPLICANT: Loosmore, Sheena M

; APPLICANT: Du, Run-Pan

; APPLICANT: Wang, Quljun

; APPLICANT: Yang, Yan-Ping

; APPLICANT: Klein, Michel H

; TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA

NUMBER OF SEQUENCES: 67

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sim & McBurney

STREET: 6th Floor, 330 University Avenue

CITY: Toronto

STATE: Ontario

COUNTRY: Canada

ZIP: M5G 1R7

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/867,941

; FILING DATE: 03-JUN-1997

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Stewart, Michael I

; REGISTRATION NUMBER: 24,973

; REFERENCE/DOCKET NUMBER: 1038-681 MIS:jb

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (416) 595-1155

; TELEFAX: (416) 595-1163

; INFORMATION FOR SEQ ID NO: 13:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 985 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-867-941-13

Query Match 90.0%; Score 27; DB 2; Length 985;

Best Local Similarity 83.3%; Pred. No. 1.4e+03;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSLSGS 6  
1:||||

Db 175 GALSGS 180

RESULT 25

US-08-867-941-17

; Sequence 17, Application US/08867941  
; Patent No. 5977337

; GENERAL INFORMATION:

; APPLICANT: Loosmore, Sheena M

; APPLICANT: Du, Run-Pan

; APPLICANT: Wang, Quljun

; APPLICANT: Yang, Yan-Ping

; APPLICANT: Klein, Michel H

; TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA

; NUMBER OF SEQUENCES: 67

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sim & McBurney

; STREET: 6th Floor, 330 University Avenue

CITY: Toronto

STATE: Ontario

COUNTRY: Canada

ZIP: M5G 1R7

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/867,941

; FILING DATE: 03-JUN-1997

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Stewart, Michael I

; REGISTRATION NUMBER: 24,973



REFERENCE/DOCKET NUMBER: 1038-681 MIS:jb  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 985 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-867-941-17

Query Match 90.0%; Score 27; DB 4; Length 985;  
Best Local Similarity 83.3%; Pred. No. 1.4e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GSIGS 6  
Db 175 GALGGS 180

RESULT 26  
US-09-074-658-13  
Sequence 13, Application US/09074658  
Patent No. 6184371  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena M  
APPLICANT: Run-Pan Du  
APPLICANT: Quijun Wang  
APPLICANT: Yang, Yan-Ping  
APPLICANT: Klein, Michel H  
TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA  
NUMBER OF SEQUENCES: 78  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Slim & McBurney  
STREET: 6th Floor, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/074,658  
FILING DATE: 08-MAY-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-795  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 985 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-074-658-13

Query Match 90.0%; Score 27; DB 4; Length 985;  
Best Local Similarity 83.3%; Pred. No. 1.4e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GSIGS 6  
Db 175 GALGGS 180

RESULT 27  
US-09-074-658-17  
Sequence 17, Application US/09074658  
Patent No. 6184371  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena M  
APPLICANT: Run-Pan Du  
APPLICANT: Quijun Wang  
APPLICANT: Yang, Yan-Ping  
APPLICANT: Klein, Michel H  
TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA  
NUMBER OF SEQUENCES: 78  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Slim & McBurney  
STREET: 6th Floor, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/074,658  
FILING DATE: 08-MAY-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-795  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 985 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-074-658-17

Query Match 90.0%; Score 27; DB 4; Length 985;  
Best Local Similarity 83.3%; Pred. No. 1.4e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GSIGS 6  
Db 175 GALGGS 180

RESULT 28  
US-08-867-941-12  
Sequence 12, Application US/08867941  
Patent No. 5977337  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena M  
APPLICANT: Du, Run-Pan  
APPLICANT: Wang, Quijun  
APPLICANT: Yang, Yan-Ping  
APPLICANT: Klein, Michel H  
TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA  
NUMBER OF SEQUENCES: 67  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Slim & McBurney  
STREET: 6th Floor, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada

```

; ZIP: MSG 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/867,941
; FILING DATE: 03-JUN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-681 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1000 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-867-941-12

Query Match          90.0%; Score 27; DB 2; Length 1000;
Best Local Similarity 83.3%; Pred. No. 1.4e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSIGS 6
DB 190 GALGS 195

RESULT 29
US-08-867-941-16
; Sequence 16, Application US/08867941
; Patent No. 5977337
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M
; APPLICANT: Du, Run-Pan
; APPLICANT: Wang, Quijun
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Slim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: MSG 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/867,941
; FILING DATE: 03-JUN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-681 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1000 amino acids
```

```

; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-867-941-16

Query Match          90.0%; Score 27; DB 2; Length 1000;
Best Local Similarity 83.3%; Pred. No. 1.4e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSIGS 6
DB 190 GALGS 195

RESULT 30
US-09-074-658-12
; Sequence 12, Application US/09074658
; Patent No. 6184371
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M
; APPLICANT: Run-Pan Du
; APPLICANT: Quijun Wang
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA
; NUMBER OF SEQUENCES: 78
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Slim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: MSG 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/074,658
; FILING DATE: 08-MAY-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-795
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1000 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-074-658-12

Query Match          90.0%; Score 27; DB 4; Length 1000;
Best Local Similarity 83.3%; Pred. No. 1.4e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSIGS 6
DB 190 GALGS 195

RESULT 31
US-09-074-658-16
; Sequence 16, Application US/09074658
; Patent No. 6184371
; GENERAL INFORMATION:
```

APPLICANT: Loosmore, Sheena M  
APPLICANT: Run-Pan Du  
APPLICANT: Quijun Wang  
APPLICANT: Yang, Yan-Ping  
APPLICANT: Klein, Michel H  
TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA  
NUMBER OF SEQUENCES: 78  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: 6th Floor, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/074,658  
FILING DATE: 08-MAY-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-795  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1000 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-074-658-16

Query Match 90.0%; Score 27; DB 4; Length 1000;  
Best Local Similarity 83.3%; Pred. No. 1.4e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GSLGS 6  
Db 190 GALGS 195

RESULT 32  
US-08-484-438-9  
Sequence 9, Application US/08484438  
Patent No. 5811098  
Patent No. 5811098 5780031  
GENERAL INFORMATION:  
APPLICANT: Plozman, Gregory D.  
APPLICANT: Culouscou, Jean-Michel  
APPLICANT: Shoyab, Mohammed  
APPLICANT: Siegal, Clay B.  
APPLICANT: Hellstr m, Ingegerd  
APPLICANT: Hellstr m, Karl E.  
TITLE OF INVENTION: HERA HUMAN RECEPTOR TYROSINE KINASE  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennle & Edmunds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/484,438  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/323,442  
FILING DATE: 14-OCT-1994  
APPLICATION NUMBER: US 08/150,704  
FILING DATE: 10-NOV-1993  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/981,165  
FILING DATE: 24-NOV-1992  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 5624-230  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1342 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-484-438-9

Query Match 90.0%; Score 27; DB 2; Length 1342;  
Best Local Similarity 83.3%; Pred. No. 1.9e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GSLGS 6  
Db 1060 GNLGS 1065

RESULT 33  
US-09-074-658-15  
Sequence 15, Application US/09074658  
Patent No. 6184371  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena M  
APPLICANT: Run-Pan Du  
APPLICANT: Quijun Wang  
APPLICANT: Yang, Yan-Ping  
APPLICANT: Klein, Michel H  
TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA  
NUMBER OF SEQUENCES: 78  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: 6th Floor, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/074,658  
FILING DATE: 08-MAY-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973

REFERENCE/DOCKET NUMBER: 1038-795  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2432 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-074-658-15

Query Match 90.0%; Score 27; DB 4; Length 2432;  
Best Local Similarity 83.3%; Pred. No. 3.6e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSLGGS 6  
Db 1084 GALGGS 1089

RESULT 34  
US-09-074-658-11  
Sequence 11, Application US/09074658  
Patent No. 6184371  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena M  
APPLICANT: Kun-Pan Du  
APPLICANT: Quljun Wang  
APPLICANT: Yang, Yan-Ping  
APPLICANT: Klein, Michel H  
TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA  
NUMBER OF SEQUENCES: 78  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & Mcburney  
STREET: 6th Floor, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/074,658  
FILING DATE: 08-MAY-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-795  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2439 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-074-658-11

Query Match 90.0%; Score 27; DB 4; Length 2439;  
Best Local Similarity 83.3%; Pred. No. 3.6e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSLGGS 6  
Db 1608 GALGGS 1093

RESULT 35  
US-07-609-716-16  
Sequence 16, Application US/07609716  
Patent No. 5514581  
GENERAL INFORMATION:  
APPLICANT: Ferrari, Franco A.  
APPLICANT: Cappello, Joseph  
TITLE OF INVENTION: Functional Recombinantly Prepared  
TITLE OF INVENTION: Synthetic Protein Polymer  
NUMBER OF SEQUENCES: 118  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: CA  
COUNTRY: US  
ZIP: 94111

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/609,716  
FILING DATE: 06-NOV-1990  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Rowland, Bertam I  
REGISTRATION NUMBER: 20015  
REFERENCE/DOCKET NUMBER: A-55186-3/BIR  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-781-1989  
TELEFAX: 415-398-3249  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-07-609-716-16

Query Match 86.7%; Score 26; DB 1; Length 15;  
Best Local Similarity 83.3%; Pred. No. 27;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GSLGGS 6  
Db 4 GSFSGS 9

RESULT 36  
US-07-609-716-67  
Sequence 67, Application US/07609716  
Patent No. 5514581  
GENERAL INFORMATION:  
APPLICANT: Ferrari, Franco A.  
APPLICANT: Cappello, Joseph  
TITLE OF INVENTION: Functional Recombinantly Prepared  
TITLE OF INVENTION: Synthetic Protein Polymer  
NUMBER OF SEQUENCES: 118  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: CA  
COUNTRY: US  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/609,716  
FILING DATE: 06-NOV-1990  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Rowland, Bertam I  
REGISTRATION NUMBER: 20015  
REFERENCE/DOCKET NUMBER: A-55186-3/BIR  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-781-1989  
TELEFAX: 415-398-3249  
INFORMATION FOR SEQ ID NO: 67:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-07-609-716-67

Query Match 86.7%; Score 26; DB 1; Length 15;  
Best Local Similarity 83.3%; Pred. No. 27;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GSLSGS 6  
|||  
DB 4 GSFGGS 9

RESULT 37  
US-08-477-509B-37  
Sequence 37, Application US/08477509B  
Patent No. 5770697  
GENERAL INFORMATION:  
APPLICANT: Ferrari, Franco A  
APPLICANT: Cappello, Joseph  
APPLICANT: Crisman, John W  
APPLICANT: Dorman, Mary A  
TITLE OF INVENTION: No. 5770697e1 Peptides Comprising Repetitive  
TITLE OF INVENTION: Units of Amino Acids and DNA Sequences Encoding the Same  
NUMBER OF SEQUENCES: 112  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Flehr, Hobbach, Test, Albrighton & Herbert  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: US  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/477,509B  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/175,155  
FILING DATE: 29-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/053,049  
FILING DATE: 22-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/114,618  
FILING DATE: 29-OCT-1987  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 06/927,258  
FILING DATE: 04-NOV-1986

ATTORNEY/AGENT INFORMATION:  
NAME: Treccartin, Richard F.  
REGISTRATION NUMBER: 31,801  
REFERENCE/DOCKET NUMBER: A-55186-7/RFT/MTK  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-781-1989  
TELEFAX: 415-398-3249  
INFORMATION FOR SEQ ID NO: 37:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-477-509B-37

Query Match 86.7%; Score 26; DB 1; Length 15;  
Best Local Similarity 83.3%; Pred. No. 27;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GSLSGS 6  
|||  
DB 4 GSFGGS 9

RESULT 38  
US-08-482-085B-37  
Sequence 37, Application US/08482085B  
Patent No. 6018030  
GENERAL INFORMATION:  
APPLICANT: Ferrari, Franco A.  
APPLICANT: Richardson, Charles  
APPLICANT: Chambers, James  
APPLICANT: Causey, Stuart  
APPLICANT: Pollock, Thomas J.  
APPLICANT: Cappello, Joseph  
APPLICANT: Crisman, John W  
TITLE OF INVENTION: No. 6018030e1 Peptides Comprising Repetitive  
TITLE OF INVENTION: Units of Amino Acids and DNA Sequences Encoding the Same  
NUMBER OF SEQUENCES: 112  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Flehr, Hobbach, Test, Albrighton & Herbert  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: US  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/482,085B  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 06/927,258  
FILING DATE: 04-NOV-1986  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/114,618  
FILING DATE: 29-OCT-1987  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/053,049  
FILING DATE: 22-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/175,155  
FILING DATE: 29-DEC-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Treccartin, Richard F.  
REGISTRATION NUMBER: 31,801  
REFERENCE/DOCKET NUMBER: A-55186-6/RFT/MTK

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-781-1989  
TELEFAX: 415-398-3249  
INFORMATION FOR SEQ ID NO: 37:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-482-085b-37

Query Match 86.7%; Score 26; DB 3; Length 15;  
Best Local Similarity 83.3%; Pred. No. 27;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GSLGGS 6  
|||  
DB 4 GSFGGS 9

RESULT 39  
US-08-475-411A-16  
Sequence 16, Application US/08475411A  
Patent No. 6140072  
GENERAL INFORMATION:  
APPLICANT: Ferrari, Franco A.  
APPLICANT: Cappello, Joseph  
TITLE OF INVENTION: Functional Recombinantly Prepared  
NUMBER OF INVENTION: Synthetic Protein Polymer  
NUMBER OF SEQUENCES: 119  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: CA  
COUNTRY: US  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/475,411A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/609,716  
FILING DATE: 06-NOV-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/269,429  
FILING DATE: 09-NOV-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/114,618  
FILING DATE: 29-OCT-1987  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 06/927,258  
FILING DATE: 04-NOV-1986  
ATTORNEY/AGENT INFORMATION:  
NAME: Treacartin, Richard F.  
REGISTRATION NUMBER: 31,801  
REFERENCE/DOCKET NUMBER: A-55186-9/RFT/MTK  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-781-1989  
TELEFAX: 415-398-3249  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: peptide  
US-08-475-411A-16

Query Match 86.7%; Score 26; DB 4; Length 15;  
Best Local Similarity 83.3%; Pred. No. 27;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GSLGGS 6  
|||  
DB 4 GSFGGS 9

RESULT 40  
US-08-475-411A-67  
Sequence 67, Application US/08475411A  
Patent No. 6140072  
GENERAL INFORMATION:  
APPLICANT: Ferrari, Franco A.  
APPLICANT: Cappello, Joseph  
TITLE OF INVENTION: Functional Recombinantly Prepared  
NUMBER OF INVENTION: Synthetic Protein Polymer  
NUMBER OF SEQUENCES: 119  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: CA  
COUNTRY: US  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/475,411A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/609,716  
FILING DATE: 06-NOV-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/269,429  
FILING DATE: 09-NOV-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/114,618  
FILING DATE: 29-OCT-1987  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 06/927,258  
FILING DATE: 04-NOV-1986  
ATTORNEY/AGENT INFORMATION:  
NAME: Treacartin, Richard F.  
REGISTRATION NUMBER: 31,801  
REFERENCE/DOCKET NUMBER: A-55186-9/RFT/MTK  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-781-1989  
TELEFAX: 415-398-3249  
INFORMATION FOR SEQ ID NO: 67:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-475-411A-67

Query Match 86.7%; Score 26; DB 4; Length 15;  
Best Local Similarity 83.3%; Pred. No. 27;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GSLGGS 6

11 111  
Db 4 GSFCGS 9

Search completed: February 4, 2002, 08:01:38  
Job time: 82 sec

---

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 4, 2002, 08:00:16 ; Search time 19.95 Seconds  
(without alignments)  
6.768 Million cell updates/sec

Title: US-09-642-660-10  
Perfect score: 33  
Sequence: 1 GCGTSC 6

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PTOS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/Dackfilest.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33	100.0	6	US-08-828-712-10	Sequence 10, Appl
2	33	100.0	6	US-09-063-276-10	Sequence 10, Appl
3	33	100.0	348	US-08-974-546-1	Sequence 1, Appl
4	33	100.0	2071	US-09-415-522-6	Sequence 6, Appl
5	30	90.9	32	US-08-461-990B-14	Sequence 14, Appl
6	30	90.9	95	US-08-461-990B-26	Sequence 26, Appl
7	30	90.9	537	US-09-028-934-25	Sequence 29, Appl
8	30	90.9	538	US-08-258-261B-2	Sequence 2, Appl
9	30	90.9	538	US-08-456-837-2	Sequence 2, Appl
10	30	90.9	538	US-08-457-342-2	Sequence 2, Appl
11	30	90.9	538	US-08-457-646A-2	Sequence 2, Appl
12	30	90.9	538	US-08-458-076A-2	Sequence 2, Appl
13	30	90.9	538	US-08-457-335A-2	Sequence 2, Appl
14	30	90.9	538	US-08-729-214-2	Sequence 2, Appl
15	30	90.9	538	US-08-729-214-24	Sequence 2, Appl
16	30	90.9	538	US-09-028-934-2	Sequence 2, Appl
17	30	90.9	538	US-09-028-934-24	Sequence 2, Appl
18	30	90.9	543	US-09-199-229-2	Sequence 2, Appl
19	30	90.9	543	US-09-443-087-2	Sequence 2, Appl
20	29	87.9	15	US-08-955-846A-45	Sequence 45, Appl
21	29	87.9	15	US-08-501-253A-9	Sequence 9, Appl
22	29	87.9	24	US-08-256-156A-2	Sequence 2, Appl
23	29	87.9	42	US-08-377-687-39	Sequence 39, Appl
24	29	87.9	42	US-08-777-197-39	Sequence 39, Appl
25	29	87.9	42	US-08-971-982-39	Sequence 10, Appl
26	29	87.9	71	PCT-US96-01720-10	Sequence 11, Appl
27	29	87.9	71	PCT-US96-01720-11	Sequence 11, Appl

28	29	87.9	91	US-08-325-253-10	Sequence 10, Appl
29	29	87.9	98	US-08-481-658B-50	Sequence 50, Appl
30	29	87.9	98	US-08-477-504A-50	Sequence 50, Appl
31	29	87.9	98	US-08-486-756A-50	Sequence 50, Appl
32	29	87.9	98	US-08-485-862B-50	Sequence 50, Appl
33	29	87.9	98	US-08-487-077A-50	Sequence 50, Appl
34	29	87.9	98	US-08-485-863A-50	Sequence 50, Appl
35	29	87.9	98	US-08-485-049D-50	Sequence 50, Appl
36	29	87.9	105	US-08-905-223-319	Sequence 319, Appl
37	29	87.9	106	5252466-6	Patent No. 5252466
38	29	87.9	123	US-07-956-700B-107	Sequence 107, App
39	29	87.9	123	US-08-476-537-107	Sequence 107, App
40	29	87.9	123	US-08-485-607-107	Sequence 107, App
41	29	87.9	123	US-08-475-879-107	Sequence 107, App
42	29	87.9	123	US-07-687-819-1	Sequence 1, Appl
43	29	87.9	194	US-09-364-083-2	Sequence 2, Appl
44	29	87.9	264	US-08-564-164A-4	Sequence 4, Appl
45	29	87.9	278	US-08-188-582-9	Sequence 9, Appl
46	29	87.9	278	US-08-646-715-9	Sequence 9, Appl
47	29	87.9	309	US-09-342-084-2	Sequence 2, Appl
48	29	87.9	330	US-08-712-948-2	Sequence 2, Appl
49	29	87.9	342	US-08-828-741B-6	Sequence 6, Appl
50	29	87.9	363	US-08-984-618-13	Sequence 13, Appl
51	29	87.9	363	US-08-984-618-16	Sequence 16, Appl
52	29	87.9	365	US-08-515-251A-2	Sequence 2, Appl
53	29	87.9	367	US-08-515-251A-4	Sequence 4, Appl
54	29	87.9	377	US-08-787-739-87	Sequence 87, Appl
55	29	87.9	377	US-09-178-115-87	Sequence 87, Appl
56	29	87.9	422	US-09-177-776-87	Sequence 87, Appl
57	29	87.9	422	US-08-335-469-2	Sequence 2, Appl
58	29	87.9	442	US-08-693-457-4	Sequence 2, Appl
59	29	87.9	442	US-08-693-457-4	Sequence 2, Appl
60	29	87.9	442	US-09-265-731-2	Sequence 2, Appl
61	29	87.9	442	US-09-265-731-4	Sequence 4, Appl
62	29	87.9	457	US-09-124-238A-1	Sequence 1, Appl
63	29	87.9	459	US-08-481-658B-2	Sequence 2, Appl
64	29	87.9	459	US-08-477-504A-2	Sequence 2, Appl
65	29	87.9	459	US-08-486-756A-2	Sequence 2, Appl
66	29	87.9	459	US-08-485-862B-2	Sequence 2, Appl
67	29	87.9	459	US-08-787-739-2	Sequence 2, Appl
68	29	87.9	459	US-08-487-077A-2	Sequence 2, Appl
69	29	87.9	459	US-08-485-863A-2	Sequence 2, Appl
70	29	87.9	459	US-08-485-049D-2	Sequence 2, Appl
71	29	87.9	459	US-09-178-115-2	Sequence 2, Appl
72	29	87.9	459	US-09-177-776-2	Sequence 2, Appl
73	29	87.9	495	US-08-828-741B-4	Sequence 4, Appl
74	29	87.9	499	US-08-820-170A-40	Sequence 40, Appl
75	29	87.9	499	US-09-055-699-40	Sequence 40, Appl
76	29	87.9	499	US-09-273-565-40	Sequence 40, Appl
77	29	87.9	513	US-08-464-266-2	Sequence 2, Appl
78	29	87.9	513	US-08-464-272-2	Sequence 2, Appl
79	29	87.9	513	US-08-464-514-2	Sequence 2, Appl
80	29	87.9	513	US-08-486-403-2	Sequence 2, Appl
81	29	87.9	524	US-08-624-125-2	Sequence 2, Appl
82	29	87.9	524	US-08-624-125-21	Sequence 21, Appl
83	29	87.9	540	US-08-624-125-21	Sequence 21, Appl
84	29	87.9	801	US-07-906-349A-6	Sequence 6, Appl
85	29	87.9	806	US-08-999-774A-6	Sequence 6, Appl
86	29	87.9	1046	US-08-386-727-2	Sequence 2, Appl
87	29	87.9	1046	US-08-600-452A-2	Sequence 2, Appl
88	29	87.9	1323	US-08-026-138E-4	Sequence 4, Appl
89	29	87.9	1345	US-08-977-767-3	Sequence 3, Appl
90	29	87.9	1449	US-08-938-291A-6	Sequence 6, Appl
91	29	87.9	1627	US-07-665-792E-9	Sequence 9, Appl
92	28	84.8	26	US-07-942-245-278	Sequence 278, App
93	28	84.8	26	US-07-942-245-279	Sequence 279, App
94	28	84.8	58	US-08-963-851-25	Sequence 25, Appl
95	28	84.8	84	US-08-648-323-10	Sequence 10, Appl
96	28	84.8	109	US-08-648-322-5	Sequence 5, Appl
97	28	84.8	510	US-08-220-677A-2	Sequence 2, Appl
98	28	84.8	510	US-08-246-489A-2	Sequence 2, Appl
99	28	84.8	520	US-09-030-995-3	Sequence 3, Appl
100	28	84.8	543	US-08-808-931-12	Sequence 12, Appl

## ALIGNMENTS

```
RESULT 1
US-08-828-712-10
: Sequence 10, Application US/08828712
: Patent No. 6015884
: GENERAL INFORMATION:
: APPLICANT: Schneck, Jonathan P.
: TITLE OF INVENTION: Soluble Divalent and Multivalent
: TITLE OF INVENTION: Heterodimeric Analogs of Proteins
: NUMBER OF SEQUENCES: 20
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Banner & Witcoff, Ltd.
: STREET: 1001 G Street, N.W.
: CITY: Washington
: STATE: D.C.
: COUNTRY: U.S.A.
: ZIP: 20001
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/828.712
: FILING DATE: 28-MAR-1997
: CLASSIFICATION: 514
: ATTORNEY/AGENT INFORMATION:
: NAME: Kagan, Sarah A.
: REGISTRATION NUMBER: 32,141
: REFERENCE/DOCKET NUMBER: 01107.73713
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202)508-9100
: TELEFAX: (202)508-9299
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 6 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: HYPOTHEICAL: NO
: FRAGMENT TYPE: Internal
: ORIGINAL SOURCE:
: INDIVIDUAL ISOLATE: IgG1 peptide linker
US-08-828-712-10

Query Match 100.0%; Score 33; DB 3; Length 6;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGTSG 6
Db 1 GCGTSG 6

RESULT 2
US-09-063-276-10
: Sequence 10, Application US/09063276
: Patent No. 6140113
: GENERAL INFORMATION:
: APPLICANT: Schneck, Jonathan
: APPLICANT: O'Herlin, Sean
: TITLE OF INVENTION: Molecular Complexes Which
: TITLE OF INVENTION: Modify Immune Responses
: NUMBER OF SEQUENCES: 20
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Banner & Witcoff
```

```
STREET: 1001 G Street, NW
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20001
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEQ for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/063,276
: FILING DATE: 21-APR-1998
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/828,712
: FILING DATE: 28-MAR-1997
: APPLICATION NUMBER: 60/014,367
: FILING DATE: 28-MAR-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Kagan, Sarah A
: REGISTRATION NUMBER: 32141
: REFERENCE/DOCKET NUMBER: 01107.74154
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-508-9100
: TELEFAX: 202-508-9299
: TELEX:
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 6 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
US-09-063-276-10
```

```
Query Match 100.0%; Score 33; DB 4; Length 6;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGTSG 6
Db 1 GCGTSG 6

RESULT 3
US-08-974-546-1
: Sequence 1, Application US/08974546
: Patent No. 5945287
: GENERAL INFORMATION:
: APPLICANT: Hillman, Jennifer L.
: APPLICANT: Corley, Neil C.
: APPLICANT: Shah, Purya
: TITLE OF INVENTION: TWO HUMAN HEAT SHOCK PROTEIN HOMOLOGS
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Dr.
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEQ for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/974,546
: FILING DATE: Filed Herewith
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
```

FILED DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0428  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-845-0555  
TELEFAX: 650-845-4166  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 348 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: BRAITUT21  
CLONE: 2525691  
US-08-974-546-1

Query Match 100.0%; Score 33; DB 2; Length 348;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGGTSG 6  
|||||  
DB 75 GGGTSG 80

RESULT 4  
US-09-415-522-6  
Sequence 6, Application US/09415522A  
Patent No. 6291660  
GENERAL INFORMATION:  
APPLICANT: Gaffney, Thomas  
APPLICANT: Wendland, Juergen  
APPLICANT: Philippesen, Peter  
TITLE OF INVENTION: No. 6291660e1 Fungal Genes Required For No. 6291660m1 Growth And  
FILE REFERENCE: CGC2046  
CURRENT APPLICATION NUMBER: US/09/415,522A  
CURRENT FILING DATE: 1999-10-08  
NUMBER OF SEQ ID NOS: 28  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 6  
LENGTH: 2071  
TYPE: PRT  
ORGANISM: Ashbya gossypii  
US-09-415-522-6

Query Match 100.0%; Score 33; DB 4; Length 2071;  
Best Local Similarity 100.0%; Pred. No. 9.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGGTSG 6  
|||||  
DB 199 GGGTSG 204

RESULT 5  
US-08-461-990B-14  
Sequence 14, Application US/08461990B  
Patent No. 5851810  
GENERAL INFORMATION:  
APPLICANT: JOHN S. BLANCHARD  
TITLE OF INVENTION: NUCLEIC ACID ENCODING RHODOCOCCUS  
TITLE OF INVENTION: PHENYLALANINE DEHYDROGENASE  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN  
STREET: 90 PARK AVENUE  
CITY: NEW YORK

STATE: NEW YORK  
COUNTRY: U.S.A.  
ZIP: 10016  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE DISKETTE  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/461,990B  
FILING DATE: JUNE 5, 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: CRAIG J. ARNOLD  
REGISTRATION NUMBER: 34,287  
REFERENCE/DOCKET NUMBER: 96700/370  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 697-5995  
TELEFAX: (212) 286-0854 or 286-0082  
TELEX: TWX 710-581-4766  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 32  
TYPE: AMINO ACID  
TOPOLOGY: LINEAR  
MOLECULE TYPE:  
DESCRIPTION: PROTEIN  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: B. STEAROTHEROPHILUS  
INDIVIDUAL ISOLATE: ALANINE DEHYDROGENASE  
US-08-461-990B-14

Query Match 90.9%; Score 30; DB 2; Length 32;  
Best Local Similarity 83.3%; Pred. No. 64;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGGTAG 6  
|||||  
DB 6 GGGTAG 11

RESULT 6  
US-08-461-990B-26  
Sequence 26, Application US/08461990B  
Patent No. 5851810  
GENERAL INFORMATION:  
APPLICANT: JOHN S. BLANCHARD  
TITLE OF INVENTION: NUCLEIC ACID ENCODING RHODOCOCCUS  
TITLE OF INVENTION: PHENYLALANINE DEHYDROGENASE  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN  
STREET: 90 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: U.S.A.  
ZIP: 10016  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE DISKETTE  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/461,990B  
FILING DATE: JUNE 5, 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: CRAIG J. ARNOLD  
REGISTRATION NUMBER: 34,287  
REFERENCE/DOCKET NUMBER: 96700/370  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 697-5995  
TELEFAX: (212) 286-0854 or 286-0082

TELEX: TWX 710-581-4766  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 95  
TYPE: AMINO ACID  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PROTEIN  
DESCRIPTION: NO  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: B. STEAROTHERMOPHILUS  
INDIVIDUAL ISOLATE: ALANINE DEHYDROGENASE  
US-08-461-990B-26

Query Match 90.9%; Score 30; DB 2; Length 95;  
Best Local Similarity 83.3%; Pred. No. 1.7e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGTSG 6  
||||:|  
DB 67 GCGTAG 72

RESULT 7  
US-09-028-934-29  
Sequence 29, Application US/09028934  
Patent No. 6117670  
GENERAL INFORMATION:  
APPLICANT: Ligon, James M.  
APPLICANT: Hill, Dwight S.  
APPLICANT: Lam, Steven T.  
APPLICANT: Hammer, Philip E.  
APPLICANT: Van Pee, Karl-Heinz  
APPLICANT: Kirner, Sabine  
APPLICANT: Young, Thomas R.  
TITLE OF INVENTION: Pyrroliditrin Biosynthesis Genes and Uses  
TITLE OF INVENTION: Thereof  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6117670artis Corporation  
STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: NC  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/028,934  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/729,214  
FILING DATE: 09-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/258,261  
FILING DATE: 08-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: CGC1506/C1P7  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8687  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 537 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

MOLECULE TYPE: Protein  
US-09-028-934-29

Query Match 90.9%; Score 30; DB 3; Length 537;  
Best Local Similarity 83.3%; Pred. No. 7.9e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGTSG 6  
||||:|  
DB 12 GCGTAG 17

RESULT 8  
US-08-258-261B-2  
Sequence 2, Application US/08258261B  
Patent No. 5639949  
GENERAL INFORMATION:  
APPLICANT: Schupp, Thomas  
APPLICANT: Ligon, James M.  
APPLICANT: Beck, James Joseph  
APPLICANT: Hill, Dwight Steven  
APPLICANT: Ryals, John Andrew  
APPLICANT: Gaffney, Thomas Deane  
APPLICANT: Lam, Stephen Ting  
APPLICANT: Hammer, Phillip E.  
APPLICANT: Uknes, Scott Joseph  
TITLE OF INVENTION: Genes for the synthesis of  
TITLE OF INVENTION: antipathogenic substances  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ciba-Geigy Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: NY  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/258,261B  
FILING DATE: 08-JUN-1994  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/457,205  
FILING DATE: 01-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: CGC 1506/C1P3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8614  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 538 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-258-261B-2

Query Match 90.9%; Score 30; DB 1; Length 538;  
Best Local Similarity 83.3%; Pred. No. 7.9e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGTSG 6  
||||:|  
DB 12 GCGTAG 17

```
RESULT 9
US-08-456-837-2
; Sequence 2, Application US/08456837
; Patent No. 5643774
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/456,837
; APPLICATION NUMBER: US/08/456,837
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-Jun-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 538 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-456-837-2

Query Match          90.9%: Score 30; DB 1; Length 538;
Best Local Similarity 83.3%: Pred. No. 7.9e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGTSG 6
    ||||:|
Db 12 GCGTAG 17

RESULT 10
US-08-457-342-2
; Sequence 2, Application US/08457342
; Patent No. 5662898
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
```

```
APPLICANT: Ryals, John Andrew
APPLICANT: Gaffney, Thomas Deane
APPLICANT: Lam, Stephen Ting
APPLICANT: Hammer, Phillip E.
APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/457,342
; APPLICATION NUMBER: US/08/457,342
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-Jun-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 538 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-457-342-2

Query Match          90.9%: Score 30; DB 1; Length 538;
Best Local Similarity 83.3%: Pred. No. 7.9e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGTSG 6
    ||||:|
Db 12 GCGTAG 17

RESULT 11
US-08-457-646A-2
; Sequence 2, Application US/08457646A
; Patent No. 5679560
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
```

```

: STREET: 7 Skyline Drive
: CITY: Hawthorne
: STATE: NY
: COUNTRY: USA
: ZIP: 10532
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/457,646A
: FILING DATE: 01-JUN-1995
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/457,205
: FILING DATE: 01-JUN-1995
: APPLICATION NUMBER: 08/258,261
: FILING DATE: 08-Jun-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Elmer, James Scott
: REGISTRATION NUMBER: 36,129
: REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8614
: TELEFAX: 919-541-8689
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 538 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-457-646A-2

Query Match          90.9%; Score 30; DB 1; Length 538;
Best Local Similarity 83.3%; Pred. No. 7.9e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGTSG 6
   ||||:|
Db 12 GCGTAG 17

RESULT 12
US-08-458-076A-2
: Sequence 2, Application US/08458076A
: Patent No. 5698425
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James M.
: APPLICANT: Beck, James Joseph
: APPLICANT: Hill, Dwight Steven
: APPLICANT: Ryals, John Andrew
: APPLICANT: Gaffney, Thomas Deane
: APPLICANT: Lam, Stephen Ting
: APPLICANT: Hammer, Phillip E.
: APPLICANT: Uknes, Scott Joseph
: TITLE OF INVENTION: Genes for the synthesis of
: NUMBER OF SEQUENCES: 22
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Ciba-Geigy Corporation
: STREET: 7 Skyline Drive
: CITY: Hawthorne
: STATE: NY
: COUNTRY: USA
: ZIP: 10532
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/457,646A
: FILING DATE: 01-JUN-1995
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/457,205
: FILING DATE: 01-JUN-1995
: APPLICATION NUMBER: 08/258,261
: FILING DATE: 08-Jun-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Elmer, James Scott
: REGISTRATION NUMBER: 36,129
: REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8614
: TELEFAX: 919-541-8689
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 538 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-458-076A-2
```

```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/458,076A
: FILING DATE: 01-JUN-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/457,205
: FILING DATE: 01-JUN-1995
: APPLICATION NUMBER: 08/258,261
: FILING DATE: 08-Jun-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Elmer, James Scott
: REGISTRATION NUMBER: 36,129
: REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8614
: TELEFAX: 919-541-8689
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 538 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-458-076A-2

Query Match          90.9%; Score 30; DB 1; Length 538;
Best Local Similarity 83.3%; Pred. No. 7.9e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGTSG 6
   ||||:|
Db 12 GCGTAG 17

RESULT 13
US-08-457-335A-2
: Sequence 2, Application US/08457335A
: Patent No. 5723759
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James M.
: APPLICANT: Beck, James Joseph
: APPLICANT: Hill, Dwight Steven
: APPLICANT: Ryals, John Andrew
: APPLICANT: Gaffney, Thomas Deane
: APPLICANT: Lam, Stephen Ting
: APPLICANT: Hammer, Phillip E.
: APPLICANT: Uknes, Scott Joseph
: TITLE OF INVENTION: Genes for the synthesis of
: NUMBER OF SEQUENCES: 22
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Ciba-Geigy Corporation
: STREET: 7 Skyline Drive
: CITY: Hawthorne
: STATE: NY
: COUNTRY: USA
: ZIP: 10532
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/457,335A
: FILING DATE: 01-JUN-1995
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/457,205
: FILING DATE: 01-JUN-1995
: APPLICATION NUMBER: 08/258,261
: FILING DATE: 08-Jun-1994
: ATTORNEY/AGENT INFORMATION:
```

NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: CGC 1506/CIP3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8614  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 538 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-457-335A-2

Query Match 90.9%; Score 30; DB 1; Length 538;  
Best Local Similarity 83.3%; Pred. No. 7.9e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGTSG 6  
||||:|  
Db 12 GCGTAG 17

RESULT 14  
US-08-729-214-2  
Sequence 2, Application US/08729214  
Patent No. 5817502  
GENERAL INFORMATION:  
APPLICANT: Ligon, James M.  
APPLICANT: Hill, Dwight Steven  
APPLICANT: Ryals, John Andrew  
APPLICANT: Hammer, Phillip E.  
APPLICANT: van Pee, Karl-Heinz  
APPLICANT: Kirner, Sabine  
TITLE OF INVENTION: Genes for the synthesis of  
TITLE OF INVENTION: antipathogenic substances  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ciba-Geigy Corporation  
STREET: 520 White Plains Road  
CITY: Tarrytown  
STATE: NY  
COUNTRY: USA  
ZIP: 10591  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/729,214  
FILING DATE: TBA  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: CGC 1506/CIP5  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 538 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-729-214-2

Query Match 90.9%; Score 30; DB 2; Length 538;  
Best Local Similarity 83.3%; Pred. No. 7.9e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGTSG 6  
||||:|  
Db 12 GCGTAG 17

RESULT 15  
US-08-729-214-24  
Sequence 24, Application US/08729214  
Patent No. 5817502  
GENERAL INFORMATION:  
APPLICANT: Ligon, James M.  
APPLICANT: Hill, Dwight Steven  
APPLICANT: Ryals, John Andrew  
APPLICANT: Hammer, Phillip E.  
APPLICANT: van Pee, Karl-Heinz  
APPLICANT: Kirner, Sabine  
TITLE OF INVENTION: Genes for the synthesis of  
TITLE OF INVENTION: antipathogenic substances  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ciba-Geigy Corporation  
STREET: 520 White Plains Road  
CITY: Tarrytown  
STATE: NY  
COUNTRY: USA  
ZIP: 10591  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/729,214  
FILING DATE: TBA  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: CGC 1506/CIP5  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 538 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-729-214-24

Query Match 90.9%; Score 30; DB 2; Length 538;  
Best Local Similarity 83.3%; Pred. No. 7.9e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGTSG 6  
||||:|  
Db 12 GCGTAG 17

RESULT 16  
US-09-028-934-2  
Sequence 2, Application US/09028934  
Patent No. 6117670  
GENERAL INFORMATION:  
APPLICANT: Ligon, James M.  
APPLICANT: Hill, Dwight S.  
APPLICANT: Lam, Steven T.  
APPLICANT: Hammer, Phillip E.  
APPLICANT: van Pee, Karl-Heinz  
APPLICANT: Kirner, Sabine  
APPLICANT: Young, Thomas R.

```

; TITLE OF INVENTION: Pyrolytic Biosynthesis Genes and Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6117670artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/028,934
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/729,214
; FILING DATE: 09-OCT-1996
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/258,261
; FILING DATE: 08-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CGC1506/CIP7
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 538 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-028-934-2

Query Match          90.9%; Score 30; DB 3; Length 538;
Best Local Similarity 83.3%; Pred. No. 7.9e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGTSG 6
    ||||:1
DB 12 GCGTAG 17

RESULT 17
US-09-028-934-24
; Sequence 24, Application US/09028934
; Patent No. 6117670
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M.
; APPLICANT: Hill, Dwight S.
; APPLICANT: Lam, Steven T.
; APPLICANT: Hammer, Philip E.
; APPLICANT: van Pee, Karl-Heinz
; APPLICANT: Kirner, Sabine
; APPLICANT: Young, Thomas R.
; TITLE OF INVENTION: Pyrolytic Biosynthesis Genes and Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6117670artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/028,934
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/729,214
; FILING DATE: 09-OCT-1996
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/258,261
; FILING DATE: 08-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CGC1506/CIP7
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 538 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-028-934-24
```

```

Query Match          90.9%; Score 30; DB 3; Length 538;
Best Local Similarity 83.3%; Pred. No. 7.9e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GCGTSG 6
    ||||:1
DB 12 GCGTAG 17
```

```

RESULT 18
US-09-199-229-2
; Sequence 2, Application US/09199229
; Patent No. 6063607
; GENERAL INFORMATION:
; APPLICANT: Yaver, Debbie S.
; APPLICANT: Berka, Randy W.
; APPLICANT: Rey, Michael W.
; TITLE OF INVENTION: Polypeptides Having Choline Oxidase
; TITLE OF INVENTION: Activity and Nucleic Acids Encoding Same
; FILE REFERENCE: 5735,000-US
; CURRENT APPLICATION NUMBER: US/09/199,229
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 543
; TYPE: PRT
; ORGANISM: Fusarium
; US-09-199-229-2
```

```

Query Match          90.9%; Score 30; DB 3; Length 543;
Best Local Similarity 83.3%; Pred. No. 8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GCGTSG 6
    ||||:1
DB 20 GCGTAG 25
```

```

RESULT 19
US-09-443-087-2
; Sequence 2, Application US/09443087
```



Patent No. 614664  
GENERAL INFORMATION:  
APPLICANT: Debbie Yaver  
APPLICANT: Randy M. Berka  
APPLICANT: Michael W. Rey  
TITLE OF INVENTION: Polypeptides Having Choline Oxidase  
TITLE OF INVENTION: Activity And Nucleic Acids Encoding Same  
FILE REFERENCE: 5735.200-US  
CURRENT APPLICATION NUMBER: US/09/443,087  
CURRENT FILING DATE: 1999-11-18  
EARLIER APPLICATION NUMBER: 09/199,229  
EARLIER FILING DATE: 1998-11-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 543  
TYPE: PRT  
ORGANISM: Fusarium venenatum  
US-09-443-087-2

Query Match 90.9%; Score 30; DB 4; Length 543;  
Best Local Similarity 83.3%; Pred. No. 8e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGTSG 6  
|||||  
DB 20 GCGTAG 25

RESULT 20  
US-08-955-848A-45  
Sequence 45, Application US/08955848A  
Patent No. 5969105  
GENERAL INFORMATION:  
APPLICANT: Mc Wherter, Charles  
APPLICANT: Feng, Yiding  
TITLE OF INVENTION: No. 5969105el Stem Cell Factor Receptor  
TITLE OF INVENTION: Agonists  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: G. D. Searle & Co.  
STREET: P.O. Box 5110  
CITY: Chicago  
STATE: IL  
COUNTRY: U. S. A.  
ZIP: 60680  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/955,848A  
FILING DATE: 21-OCT-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/029,165  
FILING DATE: 25-OCT-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Bennett, Dennis A  
REGISTRATION NUMBER: 34,547  
REFERENCE/DOCKET NUMBER: C-2992/1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 314-737-6986  
TELEFAX: 314-737-6972  
TELEX:  
INFORMATION FOR SEQ ID NO: 45:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

US-08-955-848A-45

Query Match 87.9%; Score 29; DB 2; Length 15;  
Best Local Similarity 83.3%; Pred. No. 47;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCGTSG 6  
|||||  
DB 5 GCGTGG 10

RESULT 21  
US-08-501-253A-9  
Sequence 9, Application US/08501253A  
Patent No. 6146628  
GENERAL INFORMATION:  
APPLICANT: Tucker, Faith  
APPLICANT: Turner, Nilgun  
TITLE OF INVENTION: Biotherapeutic Agents Comprising  
TITLE OF INVENTION: Recombinant PAP and PAP Mutants  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Merchant & Gould  
STREET: 90 South 7th Street, 3100 No. 6146628west Center  
CITY: Minneapolis  
STATE: MN  
COUNTRY: US  
ZIP: 55402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/501,253A  
FILING DATE: 11-JUL-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Kettlerberger, Denise M.  
REGISTRATION NUMBER: 33,924  
REFERENCE/DOCKET NUMBER: 600,323US01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612-332-5300  
TELEFAX: 612-332-9081  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-501-253A-9

Query Match 87.9%; Score 29; DB 4; Length 15;  
Best Local Similarity 83.3%; Pred. No. 47;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCGTSG 6  
|||||  
DB 7 GCGTGG 12

RESULT 22  
US-08-256-156A-2  
Sequence 2, Application US/08256156A  
Patent No. 5837821  
GENERAL INFORMATION:  
APPLICANT: Wu, Anna M.  
TITLE OF INVENTION: No. 5837821el Antibody Constructs  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.

STREET: Suite 701 East Tower, 555 Thirteenth  
STREET: Street, N.W.  
CITY: Washington, D.C.  
STATE: none  
COUNTRY: United States  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/256.156A  
FILING DATE: 24-JUN-1994  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Ernst, Barbara G.  
REGISTRATION NUMBER: 30,377  
REFERENCE/DOCKET NUMBER: 2124-110  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-783-6040  
TELEFAX: 202-783-6031  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 24 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-256-156A-2

Query Match 87.9%; Score 29; DB 2; Length 24;  
Best Local Similarity 83.3%; Pred. No. 71;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
OY 1 GCGTSG 6  
DB 14 GCGSSG 19

RESULT 23  
US-08-377-687-39  
Sequence 39, Application US/08377687  
Patent No. 5538525  
GENERAL INFORMATION:  
APPLICANT: BROEKAERT, WILLEM F.  
APPLICANT: CAMMUE, BRUNO P.A.  
APPLICANT: OSBORN, RUPERT W.  
APPLICANT: REES, SARAH B.  
APPLICANT: TERRAS, FRANKY R.G.  
APPLICANT: VANDERLEYDEN, JOZEF  
TITLE OF INVENTION: BIOCIDAL PROTEINS  
NUMBER OF SEQUENCES: 59  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CUSHMAN DARBY & CUSHMAN  
STREET: 1100 NEW YORK AVENUE, N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/377.687  
FILING DATE:  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/002.480  
FILING DATE: 04-JAN-1993

ATTORNEY/AGENT INFORMATION:  
NAME: KOKULIS, PAUL N.  
REGISTRATION NUMBER: 16,773  
REFERENCE/DOCKET NUMBER: 99042/SEE.36525/US/A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-861-3000  
TELEFAX: 202-822-0944  
INFORMATION FOR SEQ ID NO: 39:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 42 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-377-687-39

Query Match 87.9%; Score 29; DB 1; Length 42;  
Best Local Similarity 83.3%; Pred. No. 1.2e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
OY 1 GCGTSG 6  
DB 24 GCGSSG 29

RESULT 24  
US-08-777-192-39  
Sequence 39, Application US/08777192  
Patent No. 5824869  
GENERAL INFORMATION:  
APPLICANT: BROEKAERT, WILLEM F.  
APPLICANT: CAMMUE, BRUNO P.A.  
APPLICANT: OSBORN, RUPERT W.  
APPLICANT: REES, SARAH B.  
APPLICANT: TERRAS, FRANKY R.G.  
APPLICANT: VANDERLEYDEN, JOZEF  
TITLE OF INVENTION: BIOCIDAL PROTEINS  
NUMBER OF SEQUENCES: 59  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CUSHMAN DARBY & CUSHMAN  
STREET: 1100 NEW YORK AVENUE, N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/777.192  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/002.480  
FILING DATE: 04-JAN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: KOKULIS, PAUL N.  
REGISTRATION NUMBER: 16,773  
REFERENCE/DOCKET NUMBER: 99042/SEE.36525/US/A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-861-3000  
TELEFAX: 202-822-0944  
INFORMATION FOR SEQ ID NO: 39:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 42 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-777-192-39

Query Match 87.9%; Score 29; DB 2; Length 42;  
Best Local Similarity 83.3%; Pred. No. 1.2e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGGTSG 6  
|||||  
DB 24 GGGSSG 29

## RESULT 25

US-08-971-982-39  
; Sequence 39, Application US/08971982  
; Patent No. 6187904  
; GENERAL INFORMATION:

APPLICANT: BROKERT, WILLEM F.  
CAMMUE, BRUNO P.A.  
OSBORN, RUPERT W.  
REES, SARAH B.  
TERRAS, FRANKY R.G.  
VANDERLEIJEN, JOZEF

TITLE OF INVENTION: BIOCIDAL PROTEINS  
NUMBER OF SEQUENCES: 59

CORRESPONDENCE ADDRESS:  
ADDRESS: CUSHMAN DARBY & CUSHMAN  
STREET: 1100 NEW YORK AVENUE, N.W.  
CITY: WASHINGTON  
STATE: D.C.

COUNTRY: USA  
ZIP: 20005

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/971,982  
FILING DATE: 17-NO. 6187904-1997  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/002,480  
FILING DATE: 04-JAN-1993

ATTORNEY/AGENT INFORMATION:  
NAME: KOKULIS, PAUL N.  
REGISTRATION NUMBER: 16,773

REFERENCE/DOCKET NUMBER: 99042/SEE.36525/US/A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-861-3000  
TELEFAX: 202-822-0944

INFORMATION FOR SEQ ID NO: 39:  
SEQUENCE CHARACTERISTICS:

LENGTH: 42 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 39:

US-08-971-982-39

Query Match 87.9%; Score 29; DB 4; Length 42;  
Best Local Similarity 83.3%; Pred. No. 1.2e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGGTSG 6  
|||||  
DB 24 GGGSSG 29

RESULT 26  
PCT-US96-01720-10  
; Sequence 10, Application PC/TUS9601720

GENERAL INFORMATION:

APPLICANT:  
TITLE OF INVENTION: MODIFIED-AFFINITY STREPTAVIDIN  
NUMBER OF SEQUENCES: 11  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/01720  
FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/387,055

FILING DATE: 09-FEB-1995

ATTORNEY/AGENT INFORMATION:  
NAME: Parmelee, Steven W.  
REGISTRATION NUMBER: 31,990

REFERENCE/DOCKET NUMBER: 16336-5PC

INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:

LENGTH: 71 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: protein  
PCT-US96-01720-10

Query Match 87.9%; Score 29; DB 5; Length 71;  
Best Local Similarity 83.3%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGGTSG 6  
|||||  
DB 59 GGGTSG 64

RESULT 27  
PCT-US96-01720-11  
; Sequence 11, Application PC/TUS9601720  
; GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: MODIFIED-AFFINITY STREPTAVIDIN  
NUMBER OF SEQUENCES: 11  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/01720

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/387,055

FILING DATE: 09-FEB-1995

ATTORNEY/AGENT INFORMATION:  
NAME: Parmelee, Steven W.  
REGISTRATION NUMBER: 31,990

REFERENCE/DOCKET NUMBER: 16336-5PC

INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:

LENGTH: 71 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: protein  
PCT-US96-01720-11

Query Match 87.9%; Score 29; DB 5; Length 71;

Best Local Similarity 83.3%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCGTSG 6  
|||||  
Db 59 GCGTGG 64

RESULT 28  
US-08-325-253-10

; Sequence 10, Application US/08325253  
; Patent No. 5686564  
; GENERAL INFORMATION:  
; APPLICANT: Derek Edward Brundish  
; APPLICANT: Hans Rink  
; APPLICANT: Markus Gr(etter  
; APPLICANT: John Peter Priestle  
; TITLE OF INVENTION: Peptide Derivatives Corresponding to  
; TITLE OF INVENTION: the Carboxy Terminal Sequence of Htrudin  
; Patent No. 5686564  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CIBA-GEIGY Corporation  
; STREET: P.O. Box 12257  
; CITY: Research Triangle Park  
; STATE: No. 5686564th Carolina  
; COUNTRY: US  
; ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/325.253

FILING DATE: 20-OCT-1994

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/EP 93/00908

FILING DATE: 15-April-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: England 9209032.3

FILING DATE: 25-April-1992

ATTORNEY/AGENT INFORMATION:

NAME: W. Murray Spruill

REGISTRATION NUMBER: 32.943

REFERENCE/DOCKET NUMBER: 4-18956/A

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 91 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-325-253-10

Query Match 87.9%; Score 29; DB 1; Length 91;  
Best Local Similarity 83.3%; Pred. No. 2.3e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCGTSG 6  
|||||  
Db 71 GCGTGG 76

RESULT 29  
US-08-481-658B-50

; Sequence 50, Application US/08481658B  
; Patent No. 5955075  
; GENERAL INFORMATION:  
; APPLICANT: Zavada, Jan  
; APPLICANT: Pastorekova, Silvia

; APPLICANT: Pastorek, Jaromir  
; TITLE OF INVENTION: MN Gene and Protein  
; NUMBER OF SEQUENCES: 86  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leona L. Lauder  
; STREET: 6 Mariposa Court  
; CITY: Tiburon  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94920

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/481.658B

FILING DATE: 07-JUN-1995

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/260,190

FILING DATE: 15-JUN-1994

ATTORNEY/AGENT INFORMATION:

NAME: Lauder, Leona L.

REGISTRATION NUMBER: 30,863

REFERENCE/DOCKET NUMBER: D-0021.3E

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-435-2034

TELEFAX: 415-435-0727

INFORMATION FOR SEQ ID NO: 50:

SEQUENCE CHARACTERISTICS:

LENGTH: 98 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

DESCRIPTION: Region of homology to collagen alpha

DESCRIPTION: 1 chain

US-08-481-658B-50

Query Match 87.9%; Score 29; DB 2; Length 98;  
Best Local Similarity 83.3%; Pred. No. 2.5e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGTSG 6  
|||||  
Db 13 GCGSSG 18

RESULT 30  
US-08-477-504A-50

; Sequence 50, Application US/08477504A

; Patent No. 5972353

GENERAL INFORMATION:

APPLICANT: Pastorekova, Silvia

APPLICANT: Pastorek, Jaromir

TITLE OF INVENTION: MN Gene and Protein

NUMBER OF SEQUENCES: 86

CORRESPONDENCE ADDRESS:

ADDRESSEE: Leona L. Lauder

STREET: 6 Mariposa Court

CITY: Tiburon

STATE: California

COUNTRY: USA

ZIP: 94920

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/477.504A

FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3D  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
INFORMATION FOR SEQ ID NO: 50:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 98 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
DESCRIPTION: Region of homology to collagen alpha  
US-08-477-504A-50

Query Match 87.9%; Score 29; DB 2; Length 98;  
Best Local Similarity 83.3%; Pred. No. 2.5e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGTSG 6  
111:11  
DB 13 GGGSSG 18

RESULT 31  
US-08-486-756A-50  
Sequence 50, Application US/08486756A  
Patent No. 5981711  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
STREET: 6 Mariposa Court  
CITY: Tiburon  
STATE: California  
COUNTRY: USA  
ZIP: 94920  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/486,756A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
INFORMATION FOR SEQ ID NO: 50:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 98 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

MOLECULE TYPE: protein  
DESCRIPTION: Region of homology to collagen alpha  
US-08-486-756A-50

Query Match 87.9%; Score 29; DB 2; Length 98;  
Best Local Similarity 83.3%; Pred. No. 2.5e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGTSG 6  
111:11  
DB 13 GGGSSG 18

RESULT 32  
US-08-485-862B-50  
Sequence 50, Application US/08485862B  
Patent No. 5989838  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
STREET: 6 Mariposa Court  
CITY: Tiburon  
STATE: California  
COUNTRY: USA  
ZIP: 94920  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,862B  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/477,504  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3D  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
INFORMATION FOR SEQ ID NO: 50:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 98 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
DESCRIPTION: Region of homology to collagen alpha  
US-08-485-862B-50

Query Match 87.9%; Score 29; DB 2; Length 98;  
Best Local Similarity 83.3%; Pred. No. 2.5e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGTSG 6  
111:11  
DB 13 GGGSSG 18

RESULT 33  
US-08-487-077A-50  
Sequence 50, Application US/08487077A  
Patent No. 6069242  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
STREET: 6 Mariposa Court  
CITY: Tiburon  
STATE: California  
COUNTRY: USA  
ZIP: 94920  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,077A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3H  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
INFORMATION FOR SEO ID NO: 50:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 98 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
DESCRIPTION: Region of homology to collagen alpha  
US-08-487-077A-50

Query Match 87.9%; Score 29; DB 3; Length 98;  
Best Local Similarity 83.3%; Pred. No. 2.5e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
|||:|  
Db 13 GCGSSG 18

RESULT 34  
US-08-485-863A-50  
Sequence 50, Application US/08485863A  
Patent No. 6093548  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
STREET: 6 Mariposa Court  
CITY: Tiburon  
STATE: California  
COUNTRY: USA  
ZIP: 94920

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,863A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3G  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
INFORMATION FOR SEO ID NO: 50:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 98 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
DESCRIPTION: Region of homology to collagen alpha  
US-08-485-863A-50

Query Match 87.9%; Score 29; DB 3; Length 98;  
Best Local Similarity 83.3%; Pred. No. 2.5e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
|||:|  
Db 13 GCGSSG 18

RESULT 35  
US-08-485-049D-50  
Sequence 50, Application US/08485049D  
Patent No. 6204370  
GENERAL INFORMATION:  
APPLICANT: Zavada, Jan  
APPLICANT: Pastorekova, Silvia  
APPLICANT: Pastorek, Jaromir  
TITLE OF INVENTION: MN Gene and Protein  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
STREET: 369 Pine Street  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,049D  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/260,190  
FILING DATE: 15-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0021.3E  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-981-2034  
TELEFAX: 415-981-0332  
INFORMATION FOR SEQ ID NO: 50:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 98 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
DESCRIPTION: Region of homology to collagen alpha  
US-08-485-049D-50

Query Match 87.9%; Score 29; DB 4; Length 98;  
Best Local Similarity 83.3%; Pred. No. 2.5e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGTSG 6  
DB 13 GGGSSG 18

RESULT 36  
US-08-905-223-319

Sequence 319, Application US/08905223  
Patent No. 6222029

GENERAL INFORMATION:

APPLICANT: Edwards, Jean-Baptiste D.

APPLICANT: Duclert, Aymeric

TITLE OF INVENTION: 5' ESTS FOR SECRETED PROTEINS

NUMBER OF SEQUENCES: 503

CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobb, Martens, Olson & Bear

STREET: 501 West Broadway

CITY: San Diego

STATE: California

COUNTRY: USA

ZIP: 92101-3505

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy Disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: Win95

SOFTWARE: Word

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/905,223

FILING DATE:

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Israel, Ned A.

REGISTRATION NUMBER: 29,655

REFERENCE/DOCKET NUMBER:

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 235-8550

TELEFAX: (619) 235-0176

INFORMATION FOR SEQ ID NO: 319:

SEQUENCE CHARACTERISTICS:

LENGTH: 105 amino acids

TYPE: AMINO ACID

TOPOLOGY: LINEAR

MOLECULE TYPE: PROTEIN

ORIGINAL SOURCE:

ORGANISM: Homo Sapiens

TISSUE TYPE: Brain

FEATURE:

NAME/KEY: sig-peptide

LOCATION: -29...-1

IDENTIFICATION METHOD: Von Heijne matrix

OTHER INFORMATION: score 5.7

OTHER INFORMATION: seq SPAPLAVAGPWA/RP

US-08-905-223-319

Query Match 87.9%; Score 29; DB 4; Length 105;  
Best Local Similarity 83.3%; Pred. No. 2.6e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGTSG 6  
DB 69 GGGTGG 74

RESULT 37  
5252466-6

Patent No. 5252466

APPLICANT: CRONAN, JOHN E.

TITLE OF INVENTION: FUSION PROTEINS HAVING A SITE FOR IN

VIVO POST-TRANSLATION MODIFICATION AND METHODS OF MAKING AND

PURIFYING THEM

NUMBER OF SEQUENCES: 24

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/525,568

FILING DATE: 18-MAY-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 354,266

SEQ ID NO: 6

FILING DATE: 19-MAY-1989

LENGTH: 106

5252466-6

Query Match 87.9%; Score 29; DB 6; Length 106;  
Best Local Similarity 83.3%; Pred. No. 2.7e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGTSG 6  
DB 15 GGGTGG 20

RESULT 38  
US-07-956-700B-107

Sequence 107, Application US/07956700B

Patent No. 5539092

GENERAL INFORMATION:

APPLICANT: Robert Haselkorn and Piotr Gornicki

TITLE OF INVENTION: Cyanobacterial and Plant Acetyl-CoA

TITLE OF INVENTION: Carboxylase

NUMBER OF SEQUENCES: 116

CORRESPONDENCE ADDRESS:

ADDRESSEE: Arnold, White & Durkee

STREET: 321 No. 5539092th Clark Street

CITY: Chicago

STATE: Illinois

COUNTRY: USA

ZIP: 60610

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy Disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: ASCII-DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/956,700B

FILING DATE: 19921002

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Thomas E. No. 5539092thrup

REGISTRATION NUMBER: 33,268

REFERENCE/DOCKET NUMBER: ARCD:058

TELECOMMUNICATION INFORMATION:

TELEPHONE: 1-312-744-0090

TELEFAX: 1-312-755-4489

INFORMATION FOR SEQ ID NO: 107:

SEQUENCE CHARACTERISTICS:

LENGTH: 123 amino acids

TYPE: Amino acid

STRANDEDNESS: Single  
TOPOLOGY: Linear  
MOLECULE TYPE: Peptide  
US-07-956-700B-107

Query Match 87.9%; Score 29; DB 1; Length 123;  
Best Local Similarity 83.3%; Pred. No. 3e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
11111  
Db 32 GCGTGG 37

RESULT 39  
US-08-476-537-107  
Sequence 107, Application US/08476537  
Patent No. 5756290  
GENERAL INFORMATION:  
APPLICANT: Robert Haselkorn and Piotr Gornicki  
TITLE OF INVENTION: Cyanobacterial and Plant Acetyl-CoA  
NUMBER OF SEQUENCES: 116  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: 321 No. 5756290th Clark Street  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60610  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy Disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII-DOS  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/476,537  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/956,700  
FILING DATE: 10/21/92  
ATTORNEY/AGENT INFORMATION:  
NAME: Thomas E. No. 5756290thrup  
REGISTRATION NUMBER: 33,268  
REFERENCE/DOCKET NUMBER: ARCD:058  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 1-312-744-0090  
TELEFAX: 1-312-755-4489  
INFORMATION FOR SEQ ID NO: 107:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 123 amino acids  
TYPE: Amino acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
MOLECULE TYPE: Peptide  
US-08-476-537-107

Query Match 87.9%; Score 29; DB 1; Length 123;  
Best Local Similarity 83.3%; Pred. No. 3e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
11111  
Db 32 GCGTGG 37

RESULT 40  
US-08-485-607-107  
Sequence 107, Application US/08485607  
Patent No. 5792627

GENERAL INFORMATION:  
APPLICANT: Robert Haselkorn and Piotr Gornicki  
TITLE OF INVENTION: Cyanobacterial and Plant Acetyl-CoA  
NUMBER OF SEQUENCES: 116  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: 321 No. 5792627th Clark Street  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60610  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy Disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII-DOS  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,607  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/956,700  
FILING DATE: 10/21/92  
ATTORNEY/AGENT INFORMATION:  
NAME: Thomas E. No. 5792627thrup  
REGISTRATION NUMBER: 33,268  
REFERENCE/DOCKET NUMBER: ARCD:058  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 1-312-744-0090  
TELEFAX: 1-312-755-4489  
INFORMATION FOR SEQ ID NO: 107:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 123 amino acids  
TYPE: Amino acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
MOLECULE TYPE: Peptide  
US-08-485-607-107

Query Match 87.9%; Score 29; DB 1; Length 123;  
Best Local Similarity 83.3%; Pred. No. 3e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
11111  
Db 32 GCGTGG 37

Search completed: February 4, 2002, 08:01:36  
Job time: 80 sec





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 4, 2002, 08:00:16 ; Search time 23.49 Seconds  
(without alignments)  
19.457 Million cell updates/sec

Title: US-09-642-660-10-  
Perfect score: 33  
Sequence: 1 GCGTSG 6

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database : BIOSUM62

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	100.0	102	2 S00739	hypothetical prote
2	33	100.0	234	2 I54074	upstream stimulat
3	33	100.0	279	2 T26555	hypothetical prote
4	33	100.0	289	2 T01257	probable G1-1-like
5	33	100.0	303	2 T48493	hypothetical prote
6	33	100.0	504	2 A57215	glial cells missin
7	33	100.0	563	2 S32156	mandelonitrile lya
8	33	100.0	568	2 B23696	phosphotransferase
9	33	100.0	569	2 S32494	glucokinase regula
10	33	100.0	572	2 B28474	phosphotransferase
11	33	100.0	574	2 T50766	mandelonitrile lya
12	33	100.0	577	2 B23697	phosphotransferase
13	33	100.0	578	2 T48795	oridin recognition
14	33	100.0	582	2 F84655	hypothetical prote
15	33	100.0	625	1 S52485	glucokinase regula
16	33	100.0	627	1 S41745	glucokinase regula
17	33	100.0	629	2 A29666	keratin, 65k type
18	33	100.0	668	1 QOBEW1	UL52 protein - hum
19	33	100.0	780	2 T27941	hypothetical prote
20	33	100.0	923	1 B35905	endopeptidase Clp
21	33	100.0	926	1 A35905	endopeptidase Clp
22	33	100.0	1192	2 T18611	probable serine/th
23	33	90.9	130	2 E82699	30S ribosomal prot
24	33	90.9	136	1 S70890	hypothetical prote
25	33	90.9	136	2 D85915	hypothetical prote
26	33	90.9	143	2 T46122	30S ribosomal prot
27	33	90.9	185	2 F84182	hypothetical prote
28	33	90.9	245	2 S56827	conserved hypothet
29	33	90.9	256	2 T03371	glycine-rich prote

30	30	90.9	276	2 F75300	hypothetical prote
31	30	90.9	278	2 G85887	hypothetical prote
32	30	90.9	282	2 E65020	ethanolamine utili
33	30	90.9	310	2 JC6502	pectate lyase (EC
34	30	90.9	340	2 T20807	hypothetical prote
35	30	90.9	364	2 T34954	probable UDP-N-ace
36	30	90.9	370	2 I57555	c-Maf protein - mo
37	30	90.9	372	2 B34261	alanine dehydrogen
38	30	90.9	382	1 B46233	transcription fact
39	30	90.9	382	2 T48109	hypothetical prote
40	30	90.9	384	2 D69809	spore germination
41	30	90.9	395	2 S42680	phosphoserine tran
42	30	90.9	396	2 A71296	D-alanine--D-alani
43	30	90.9	410	2 E70579	probable murg prot
44	30	90.9	463	2 B70893	hypothetical glyci
45	30	90.9	467	2 C81138	lipeamide dehydrog
46	30	90.9	467	2 A81881	probable dihydroli
47	30	90.9	480	2 A86427	probable serine/th
48	30	90.9	484	2 E82051	pyridine nucleotid
49	30	90.9	497	2 G86299	hypothetical prote
50	30	90.9	501	2 F86275	hypothetical prote
51	30	90.9	503	2 E86275	hypothetical prote
52	30	90.9	527	2 C70397	periplasmic cell d
53	30	90.9	544	2 T44938	transducer protein
54	30	90.9	552	2 A96756	hypothetical prote
55	30	90.9	557	2 T47723	mandelonitrile lya
56	30	90.9	573	2 T07948	mandelonitrile lya
57	30	90.9	574	2 C83359	probable oxidoredu
58	30	90.9	576	2 T08073	mandelonitrile lya
59	30	90.9	589	2 T50698	probable mandeloni
60	30	90.9	594	2 T50764	adhesion of calyx
61	30	90.9	594	2 T50765	hypothetical glyci
62	30	90.9	667	2 A70893	hypothetical prote
63	30	90.9	806	2 T13690	lectin-like adhesi
64	30	90.9	819	2 A47018	hypothetical prote
65	30	90.9	896	2 T51891	dead ringer nuclea
66	30	90.9	901	2 JC6093	exonuclease ABC 3
67	30	90.9	916	2 H72372	collagen alpha 3(V
68	30	90.9	959	2 S32605	probable ATP-bindi
69	30	90.9	998	2 T35745	RNA-directed DNA p
70	30	90.9	1051	2 B27672	hypothetical prote
71	30	90.9	1108	2 A96623	hypothetical prote
72	30	90.9	1113	2 T20140	hypothetical prote
73	30	90.9	1148	2 S51855	hypothetical prote
74	30	90.9	1301	2 S18118	alpha-amyrase - Al
75	30	90.9	1314	2 T28698	hypothetical prote
76	30	90.9	3429	2 T13853	trichothax protein
77	30	90.9	3828	2 T13857	alcohol oxidase (E
78	29	87.9	33	1 A23483	flavocytochrome c
79	29	87.9	44	2 B39859	glycine rich prote
80	29	87.9	64	2 S53051	phospholipid trans
81	29	87.9	91	1 S00060	biotin carboxyl ca
82	29	87.9	123	1 BKTP	hypothetical prote
83	29	87.9	125	2 T49556	MG067 homolog D02-
84	29	87.9	135	2 S73584	hypothetical gas vesic
85	29	87.9	144	2 T34730	flavocytochrome c,
86	29	87.9	149	2 D47169	hypothetical prote
87	29	87.9	150	2 C86824	hypothetical prote
88	29	87.9	156	2 S74733	hypothetical prote
89	29	87.9	161	2 T46048	hypothetical prote
90	29	87.9	166	1 KRBO2B	keratin, 68k type
91	29	87.9	167	2 S21359	keratin, type I, c
92	29	87.9	177	2 F81442	keratin-like prote
93	29	87.9	183	2 PN0109	glycine-rich prote
94	29	87.9	201	2 F84596	glycine-rich prote
95	29	87.9	203	1 JQ1061	signal peptidase 2
96	29	87.9	226	2 A53012	transcription fact
97	29	87.9	235	2 S26300	homeotic protein U
98	29	87.9	249	2 B33144	hypothetical prote
99	29	87.9	257	2 C84890	hypothetical prote
100	29	87.9	257	2 B84346	hypothetical prote

## ALIGNMENTS

RESULT 1  
S00739  
hypothetical protein 2 - Methanococcus thermolithotrophicus (fragment)  
C:Species: Methanococcus thermolithotrophicus  
C:Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 22-Oct-1999  
C:Accession: S00739  
R:Souillard, N.; Mogol, M.; Possolet, O.; Stbold, L.  
C: Mol. Evol. 27, 65-76, 1988  
F:Title: Nucleotide sequence of regions homologous to nlfh (nitrogenase Fe protein) from  
ery implications.  
F:Reference number: S00737; MUID:88259240  
F:Accession: S00739  
F: Molecule type: DNA  
A:Residues: 1-102 <SOU>  
A:Cross-references: EMBL:X07500; NID:g44619; PIDN:CAA30382.1; PID:g44622  
A:Note: the authors translated the codon CAG for residue 91 as His

Query Match 100.0%; Score 33; DB 2; Length 102;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CY 1 GCGTSG 6  
|||||  
DB 7 GCGTSG 12

RESULT 2  
154074  
upstream stimulatory factor - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 05-Nov-1999  
C:Accession: 154074  
R:Silito, M.; Walker, S.; Lin, Q.; Kozlowski, M.T.; Klein, W.H.; Sawadogo, M.  
Gene Expr. 2, 231-240, 1992  
A:Title: Members of the USF family of helix-loop-helix proteins bind DNA as homo- as well  
A:Reference number: 154074; MUID:93082094  
A:Accession: 154074  
A:Status: preliminary; translated from GR/EMBL/DBJ  
A: Molecule type: mRNA  
A:Residues: 1-234 <RES>  
A:Cross-references: GB:S50537; NID:g261070; PIDN:AAB24368.1; PID:g261071

Query Match 100.0%; Score 33; DB 2; Length 234;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 GCGTSG 6  
|||||  
DB 1 GCGTSG 6

RESULT 3  
T26555  
hypothetical protein Y22F5A.5 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 20-Jun-2000  
C:Accession: T26555  
R:Gardner, A.  
Submitted to the EMBL Data Library, January 1998  
A:Reference number: 220231  
A:Accession: T26555  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A: Molecule type: DNA  
A:Residues: 1-279 <WIL>  
A:Cross-references: EMBL:AL021479; PIDN:CAA16324.1; GSPDB:GN00023; CESP:Y22F5A.5  
A:Experimental source: clone Y22F5A  
C:Genetics:  
A:Gene: CESP:Y22F5A.5

A:Map position: 5  
A:Introns: 142/3  
C:Superfamily: Caenorhabditis elegans hypothetical protein C02A12.4

Query Match 100.0%; Score 33; DB 2; Length 279;  
Best Local Similarity 100.0%; Pred. No. 86;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 GCGTSG 6  
|||||  
DB 206 GCGTSG 211

RESULT 4  
T01257  
probable GT-1-like transcription factor [imported] - Arabidopsis thaliana  
N:Alternate names: hypothetical protein F16M14.18  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 23-Mar-2001  
C:Accession: T01257; G84802  
R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Syke  
submitted to the EMBL Data Library, July 1998  
A:Description: Arabidopsis thaliana chromosome II BAC F16M14 genomic sequence.  
A:Reference number: Z14213  
A:Accession: T01257  
A:Status: translated from GR/EMBL/DBJ  
A: Molecule type: DNA  
A:Residues: 1-289 <ROU>  
A:Cross-references: EMBL:AC003028; NID:g3335356; PID:g3335373  
A:Experimental source: cultivar Columbia  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Sheu, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Unayam, L.; Tallon,  
Nuss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487  
A:Accession: G84802  
A:Status: preliminary  
A: Molecule type: DNA  
A:Residues: 1-289 <STO>  
A:Cross-references: GB:AE02093; NID:g3335373; PIDN:AAC27174.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: F16M14.18; AC2g38250  
A:Map position: 2  
A:Introns: 102/3

Query Match 100.0%; Score 33; DB 2; Length 289;  
Best Local Similarity 100.0%; Pred. No. 89;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 GCGTSG 6  
|||||  
DB 142 GCGTSG 147

RESULT 5  
T48493  
hypothetical protein T28J14.160 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 15-Sep-2000  
C:Accession: T48493  
R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Rudd, S.;  
submitted to the Protein Sequence Database, April 2000  
A:Reference number: 224493  
A:Accession: T48493  
A:Status: preliminary  
A: Molecule type: DNA  
A:Residues: 1-303 <BEV>  
A:Cross-references: EMBL:AL163652  
A:Experimental source: cultivar Columbia; BAC clone T28J14  
C:Genetics:

A:Map position: 5  
A:Introns: 67/1; 152/3; 201/3  
A:Note: T28J14.160  
C:Superfamily: Arabidopsis thaliana hypothetical protein P109039.6

Query Match 100.0%; Score 33; DB 2; Length 303;  
Best Local Similarity 100.0%; Pred. No. 93;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
DB 14 GCGTSG 19

RESULT 6  
A57215  
glial cells missing (gcm) protein - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C>Date: 10-Nov-1995 #sequence\_revision 10-Nov-1995 #text\_change 21-Jul-2000  
C:Accession: A57215; A57216  
R:Hosoya, T.; Takizawa, K.; Nitta, K.; Hotta, Y.  
Cell 82, 1025-1036, 1995  
A>Title: glial cells missing: a binary switch between neuronal and glial determination  
A:Reference number: A57215; MUID:96016097  
A:Accession: A57215  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-504 <CON>  
A:Cross-references: GB:U34039; NID:91041730; PIDN:AMC46912.1; PID:91041731  
C:Genetics:  
A:Gene: gcm  
A:Cross-references: FlyBase:FBgn0014179  
C:Keywords: nucleus

Query Match 100.0%; Score 33; DB 2; Length 504;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
DB 430 GCGTSG 435

RESULT 7  
S32156  
mandelonitrile lyase (EC 4.1.2.10) - black cherry  
C:Species: Prunus serotina (black cherry)  
C>Date: 22-Nov-1993 #sequence\_revision 13-Mar-1997 #text\_change 17-Mar-2000  
C:Accession: S32156  
R:Cheng, I.P.; Poulton, J.E.  
submitted to the EMBL Data Library, March 1993  
A:Description: Nucleotide sequence of a full-length cDNA clone encoding Prunus serotina  
A:Reference number: S32156  
A:Accession: S32156  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-563 <CH>  
A:Cross-references: EMBL:X72617; NID:9288115; PID:9288116  
C:Superfamily: alcohol oxidase  
C:Keywords: aldehyde-lyase; carbon-carbon lyase

Query Match 100.0%; Score 33; DB 2; Length 563;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
DB 60 GCGTSG 65

RESULT 8  
B23696  
phosphotransferase system enzyme II (EC 2.7.1.69) - Lactococcus lactis  
C:Species: Lactococcus lactis  
C>Date: 04-Oct-1991 #sequence\_revision 04-Oct-1991 #text\_change 15-Oct-1999  
C:Accession: B23696  
R:de Vos, W.M.; Boerriqter, I.; van Rooyen, R.J.; Reiche, B.; Hengstenberg, W.  
J. Biol. Chem. 265, 22554-22560, 1990  
A>Title: Characterization of the lactose-specific enzymes of the phosphotransferase s  
A:Reference number: A23696; MUID:91093107  
A:Accession: B23696  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-568 <DEV>  
A:Cross-references: GB:M60447; GB:J05748; NID:9149404; PIDN:AAA25182.1; PID:9149411  
C:Keywords: phosphotransferase; transmembrane protein

Query Match 100.0%; Score 33; DB 2; Length 568;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
DB 475 GCGTSG 480

RESULT 9  
S32494  
glucokinase regulator - rat  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 25-Dec-1994 #sequence\_revision 26-May-1995 #text\_change 17-Mar-1999  
C:Accession: S32494  
R:Delhieux, M.; Vandekerckhove, J.; van Schaftingen, E.  
FEBS Lett. 321, 111-115, 1993  
A>Title: Cloning and sequencing of rat liver cDNAs encoding the regulatory protein of  
A:Reference number: S32494; MUID:9338935  
A:Accession: S32494  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-569 <DET>  
A:Cross-references: EMBL:X68497  
A:Note: In the authors' translation residues 429-430 are shown after residue 450 and,  
C:Superfamily: glucokinase regulator

Query Match 100.0%; Score 33; DB 2; Length 569;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
DB 106 GCGTSG 111

RESULT 10  
B28474  
phosphotransferase system enzyme II (EC 2.7.1.69), lactose-specific, factor II - Stap  
C:Species: Staphylococcus aureus  
C>Date: 20-Jun-1989 #sequence\_revision 20-Jun-1989 #text\_change 15-Oct-1999  
C:Accession: B28474  
R:Bridt Jr., F.; Hengstenberg, W.; Finkeldei, U.; Stewart, G.C.  
J. Biol. Chem. 262, 16444-16449, 1987  
A>Title: Identification of the genes for the lactose-specific components of the phosp

A:Reference number: A92619; MUID:86059023  
A:Accession: B28474  
A:Molecule type: DNA  
A:Residues: 1-572 <BBR>  
A:Cross-references: GB:J03479; NID:g153036; PIDN:AAA26649.1; PID:g153038  
C:Keywords: phosphotransferase

Query Match 100.0%; Score 33; DB 2; Length 572;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Q> 1 GCGTSG 6  
|||||  
Db 478 GCGTSG 483

RESULT 11

mundionitrile lyase (EC 4.1.2.10) isoform MDL4 precursor [Imported] - black cherry  
C:Species: Prunus serotina (black cherry)  
C:Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 02-Sep-2000  
C:Accession: T50766  
R:Hu, Z.; Poulton, J.E.  
submitted to the EMBL Data Library, January 1998  
A:Reference number: Z25225  
A:Accession: T50766  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-574 <HUZ>  
A:Cross-references: EMBL:AF043186; PIDN:AAD02265.1  
C:Superfamily: alcohol oxidase  
C:Keywords: aldehyde-lyase; carbon-carbon lyase

Query Match 100.0%; Score 33; DB 2; Length 574;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Q> 1 GCGTSG 6  
|||||  
Db 61 GCGTSG 66

RESULT 12

phosphotransferase system enzyme II (EC 2.7.1.69) - Lactobacillus casei  
C:Species: Lactobacillus casei  
C:Date: 04-Oct-1991 #sequence\_revision 04-Oct-1991 #text\_change 15-Oct-1999  
C:Accession: B23697  
R:Alpert, C.A.; Chassy, B.M.  
J Biol. Chem. 265, 22561-22568, 1990  
A:Title: Molecular cloning and DNA sequence of lacE, the gene encoding the lactose-specific  
nial for sugar phosphorylation.  
A:Reference number: A23697; MUID:91093108  
A:Accession: B23697  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-577 <ALP>  
A:Cross-references: GB:M60851; NID:g149561; PIDN:AA72984.1; PID:g149563  
C:Keywords: phosphotransferase; transmembrane protein

Query Match 100.0%; Score 33; DB 2; Length 577;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Q> 1 GCGTSG 6  
|||||  
Db 485 GCGTSG 490

RESULT 13

T48795  
origin recognition complex subunit 2 related protein [Imported] - Neurospora crassa  
N:Alternate names: protein 15E6.30  
C:Species: Neurospora crassa  
C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 05-May-2000  
C:Accession: T48795  
R:Schulte, U.; Allyn, V.; Hohelsel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Myakatu  
submitted to the Protein Sequence Database, April 2000  
A:Reference number: 224541  
A:Accession: T48795  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-578 <SCH>  
A:Cross-references: EMBL:AL353822; GSPDB:GN00112; NCSP:15E6.30  
A:Experimental source: cosmid contig 15E6; strain 74  
C:Genetics:  
A:Gene: NCSP:15E6.30  
A:Map position: 2

Query Match 100.0%; Score 33; DB 2; Length 578;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Q> 1 GCGTSG 6  
|||||  
Db 521 GCGTSG 526

RESULT 14

hypothetical protein At2g42580 [Imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: F84855  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Beitel, M.T.; Town, C.D.; Fujii, C.Y.  
M.; Koo, H.; Moffet, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,  
euss, D.; Niemman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487  
A:Accession: F84855  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-618 <STO>  
A:Cross-references: GB:AE002093; NID:g4559333; PIDN:AAD22995.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g42580  
A:Map position: 2

Query Match 100.0%; Score 33; DB 2; Length 618;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Q> 1 GCGTSG 6  
|||||  
Db 203 GCGTSG 208

RESULT 15

glucokinase regulator - human  
C:Species: Homo sapiens (man)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: S52485  
R:Bonthron, D.T.; Intody, S.; Warner, J.P.  
submitted to the EMBL Data Library, February 1995  
A:Description: Human glucokinase.  
A:Reference number: S52485  
A:Accession: S52485  
A:Status: preliminary  
A:Molecule type: mRNA

A;Residues: 1-625 <BON>  
A;Cross-references: EMBL:Z48475; NID:g683571; PIDN:CAA8367.1; PID:g683572  
C;Genetics:  
A;Gene: GDB:GCKR  
A;Cross-references: GDB:207312; OMIM:600842  
A;Map position: 2p23.3-2p23.2  
C;Superfamily: glucokinase regulator

Query Match  
Best Local Similarity 100.0%; Score 33; DB 1; Length 625;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
|||||  
Db 106 GCGTSG 111

RESULT 16  
S41745  
glucokinase regulator - rat  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C;Accession: S41745  
R;Deleux, M.; Vandekerckhove, J.; van Schaftingen, E.  
FEBS Lett. 339, 312-315, 1994  
A;Title: Cloning and sequencing of rat liver cDNAs encoding the regulatory protein of gl  
A;Reference number: S41745; MUID:94156054  
A;Accession: S41745  
A;Molecule type: mRNA  
A;Residues: 1-627 <DET>  
C;Superfamily: glucokinase regulator

Query Match  
Best Local Similarity 100.0%; Score 33; DB 1; Length 627;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
|||||  
Db 106 GCGTSG 111

RESULT 17  
A29666  
keratin, 65K type II cytoskeletal - human  
C;Species: Homo sapiens (man)  
C;Date: 31-Dec-1988 #sequence\_revision 30-Sep-1991 #text\_change 20-Jun-2000  
C;Accession: A29666  
R;Klinge, E.M.; Sylvestre, Y.R.; Freedberg, I.M.; Blumenberg, M.  
J. Mol. Evol. 24, 319-329, 1987  
A;Title: Evolution of keratin genes: different protein domains evolve by different pathw  
A;Reference number: A29666; MUID:87254239  
A;Accession: A29666  
A;Molecule type: DNA  
A;Residues: 1-629 <KIL>  
A;Cross-references: GB:X05418; NID:g34040; PIDN:CAA28991.1; PID:g34041; GB:X05419; NID:g  
A28994.1; PID:g1335193; NID:g34043; PIDN:CAA28995.1; PID:g1335194; GB:X05421; NID:g34047  
A;Note: the authors translated the codon AAC for residue 63 as Asp and ACA for residue 2  
s mistranslated as a Met initiator codon  
C;Genetics:  
A;Introns: 215/3; 289/2; 309/3; 341/3; 396/3; 438/3; 512/2; 525/1  
C;Superfamily: cytoskeletal keratin  
C;Keywords: coiled coil

Query Match  
Best Local Similarity 100.0%; Score 33; DB 2; Length 629;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 GCGTSG 6  
|||||

Db 578 GCGTSG 583

RESULT 18  
Q08EM1  
US12 protein - human cytomegalovirus (strain AD169)  
N;Alternate names: HFRF1 protein  
C;Species: human cytomegalovirus, human herpesvirus 5  
A;Note: host Homo sapiens (man)  
C;Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 16-Jul-1999  
C;Accession: S09815  
R;Chen, M.S.; Bankier, A.T.; Beck, S.; Bohml, R.; Brown, C.M.; Cerny, R.; Hornsneil, T  
M.; Barrell, B.G.  
Curr. Top. Microbiol. Immunol. 154, 125-169, 1990  
A;Title: Analysis of the protein-coding content of the sequence of human cytomegalovi  
A;Reference number: S09749; MUID:90269039  
A;Accession: S09815  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-668 <CHC>  
A;Cross-references: EMBL:X17403; NID:g59591; PIDN:CAA35411.1; PID:g1780830  
A;Note: possible protein-coding frames are given  
A;Note: the DNA sequence was submitted to EMBL, December 1989, in computer-readable f  
C;Superfamily: varicella-zoster virus gene 26 protein

Query Match  
Best Local Similarity 100.0%; Score 33; DB 1; Length 668;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
|||||  
Db 422 GCGTSG 427

RESULT 19  
T27941  
hypothetical protein ZK662.3 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C;Accession: T27941  
R;White, S.  
Submitted to the EMBL Data Library, August 1996  
A;Reference number: Z20444  
A;Accession: T27941  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-780 <WIL>  
A;Cross-references: EMBL:Z79604; PIDN:CA01900.1; GSPDB:GN00028; CESP:ZK662.3  
C;Genetics:  
A;Experimental source: clone ZK662  
A;Gene: CESP:ZK662.3  
A;Map position: X  
A;Introns: 30/1; 56/1; 98/3; 176/3; 206/3; 289/2; 311/1; 341/3; 444/3; 507/3; 655/1;

Query Match  
Best Local Similarity 100.0%; Score 33; DB 2; Length 780;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
|||||  
Db 64 GCGTSG 69

RESULT 20  
B35905  
endopeptidase Clp (EC 3.4.21.-) ATP-binding chain cd4b, chloroplast [similarity] - to  
N;Alternate names: ATP-dependent Clp protease regulatory chain B; CD4B protein  
N;Contents: adenosinetriphosphatase (EC 3.6.1.3)  
C;Species: Lycopersicon esculentum (tomato)  
C;Date: 30-Jun-1993 #sequence\_revision 30-Jun-1993 #text\_change 19-Jan-2001  
C;Accession: B35905

R:Gottesman, S.; Squires, C.; Pichersky, E.; Carrington, M.; Hobbs, M.; Mattick, J.S.; D  
Proc. Natl. Acad. Sci. U.S.A. 87, 3513-3517, 1990  
A:Title: Conservation of the regulatory subunit for the Clp ATP-dependent protease in pr  
A:Reference number: A35905; MUID:90239044  
A:Accession: B35905  
A:Molecule type: DNA  
A:Residues: 1-923 <GGT>  
A:Cross-references: GB:M32604; NID:g1704324; PIDN:AAA34161.1; PID:g170435  
C:Genetics:  
A:Gene: cddA  
A:Map position: 12  
C:Function:  
A:Description: allows clpP to hydrolyze polypeptides and proteins, probably by a chaperon  
e activity: ATP hydrolysis is required for Clp hydrolysis of proteins but not of smaller  
C:Superfamily: endopeptidase Clp ATP-binding chain  
C:Keywords: ATP; chloroplast; hydrolase; molecular chaperone; nucleotide binding; P-loop  
F:300-307/Region: nucleotide-binding motif A (P-loop)  
F:367-371/Region: nucleotide-binding motif B  
F:643-650/Region: nucleotide-binding motif A (P-loop)  
F:711-715/Region: nucleotide-binding motif B  
F:306/Binding site: ATP (Lys) #status predicted  
F:649/Binding site: ATP (Lys) #status predicted

Query Match 100.0%; Score 33; DB 1; Length 923;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
|||||  
Db 244 GCGTSG 249

RESULT 21  
A35905  
Endopeptidase Clp (EC 3.4.21.-) ATP-binding chain cddA, chloroplast [similarity] - tomat  
N:Alternate names: ATP-dependent Clp proteinase regulatory chain A; CddA protein  
N:Contains: adenosinetriphosphatase (EC 3.6.1.3)  
C:Species: Lycopersicon esculentum (tomato)  
C:Date: 30-Jun-1993 #sequence\_revision 30-Jun-1993 #text\_change 19-Jan-2001  
C:Accession: A35905  
R:Gottesman, S.; Squires, C.; Pichersky, E.; Carrington, M.; Hobbs, M.; Mattick, J.S.; D  
Proc. Natl. Acad. Sci. U.S.A. 87, 3513-3517, 1990  
A:Title: Conservation of the regulatory subunit for the Clp ATP-dependent protease in pr  
A:Reference number: A35905; MUID:90239044  
A:Accession: A35905  
A:Molecule type: DNA  
A:Residues: 1-926 <GGT>  
A:Cross-references: GB:M32603; NID:g170432; PIDN:AAA34160.1; PID:g170433  
C:Genetics:  
A:Gene: cddA  
A:Map position: 3  
C:Function:  
A:Description: allows clpP to hydrolyze polypeptides and proteins, probably by a chaperon  
e activity: ATP hydrolysis is required for Clp hydrolysis of proteins but not of smaller  
C:Superfamily: endopeptidase Clp ATP-binding chain  
C:Keywords: ATP; chloroplast; hydrolase; molecular chaperone; nucleotide binding; P-loop  
F:300-309/Region: nucleotide-binding motif A (P-loop)  
F:367-371/Region: nucleotide-binding motif B  
F:646-653/Region: nucleotide-binding motif A (P-loop)  
F:711-715/Region: nucleotide-binding motif B  
F:308/Binding site: ATP (Lys) #status predicted  
F:652/Binding site: ATP (Lys) #status predicted

Query Match 100.0%; Score 33; DB 1; Length 926;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
|||||  
Db 246 GCGTSG 251

RESULT 22  
T18611  
probable serine/threonine-specific protein kinase (EC 2.7.1.-), long splice form - Ca  
N:Contains: probable serine/threonine kinase, short splice form  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 04-Mar-2000  
C:Accession: T18611; T23144; T23143  
R:McMurray, A.  
Submitted to the EMBL Data Library, October 1996  
A:Reference number: Z18997  
A:Accession: T18611  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1192 <MIL>  
A:Cross-references: EMBL:Z81027; PIDN:CAB54179.1; GSPDB:GN00023; CESP:H39E23.1a  
A:Experimental source: clone AH10  
A:Accession: T18610  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-487,536-1192 <MIL2>  
A:Cross-references: EMBL:Z81027; PIDN:CAB54178.1; GSPDB:GN00023; CESP:H39E23.1b  
A:Experimental source: clone AH10  
R:McMurray, A.  
Submitted to the EMBL Data Library, June 1997  
A:Reference number: Z19696  
A:Accession: T23144  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1192 <MIL3>  
A:Cross-references: EMBL:Z96102; PIDN:CAB54263.1; GSPDB:GN00023; CESP:H39E23.1a  
A:Experimental source: clone H39E23  
A:Accession: T23143  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-487,536-1192 <MIL4>  
A:Cross-references: EMBL:Z96102; PIDN:CAB54262.1; GSPDB:GN00023; CESP:H39E23.1b  
A:Experimental source: clone H39E23  
C:Genetics:  
A:Gene: CESP:H39E23.1a; CESP:H39E23.1b  
A:Map position: 5  
A:Initrions: 10/3; 33/3; 67/3; 139/2; 183/3; 451/3; 487/3; 535/3; 631/1; 825/2; 914/3;  
C:Keywords: Alternative splicing; ATP; phosphotransferase; serine/threonine-specific  
F:1-1192/Product: probable serine/threonine-specific protein kinase, long splice form  
F:1-487,536-1192/Product: probable serine/threonine-specific protein kinase, short sp

Query Match 100.0%; Score 33; DB 2; Length 1192;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
|||||  
Db 101 GCGTSG 106

RESULT 23  
E82669  
30S ribosomal protein S9 Xr1536 [imported] - Xylella fastidiosa (strain 9a5c)  
C:Species: Xylella fastidiosa  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000  
C:Accession: E82669  
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq  
Nature 406, 151-157, 2000  
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A:Reference number: A82515; MUID:20365717  
A:Note: for a complete list of authors see reference number A59328 below  
A:Accession: E82669  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-130 <SIM>  
A:Cross-references: GB:AE003983; GB:AE003849; NID:g9106567; PIDN:AAF84345.1; GSPDB:GN  
A:Experimental source: strain 9a5c



R.Simpson, A.J.G.; Relnach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A. Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Carraro, L.E.A.; Carraro, D.M.; Carreir, H. as-Nero, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.

A:Submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigz chado, M.A.; Madeira, A.M.B.N.; Madeira, A.M.F.; Marino, C.L.; Marques, M.V.; Martins, E A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawash A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir M.; Tsuchino, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z A:Reference number: A59328

A:Contents: annotation

A:Genetics:

C:Gene: XFL536

C:Superfamily: Escherichia coli ribosomal protein S9

Query Match 90.9%; Score 30; DB 2; Length 130;  
Best Local Similarity 83.3%; Pred. No. 1.4e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGGTSG 6  
|||||  
Db 69 GGGTTG 74

RESULT 24  
S70890  
hypothetical protein b2674 - Escherichia coli  
C:Species: Escherichia coli  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000  
C:Accession: S70890; C65047  
R:Jordan, A.; Aragall, E.; Gilbert, I.; Barbe, J.  
Mol. Microbiol. 19, 777-790, 1996  
A:Title: Promoter identification and expression analysis of Salmonella typhimurium and H A:Reference number: S70890; MUID:96417857  
A:Accession: S70890  
A:Status: preliminary; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-136 <COR>  
A:Cross-references: EMBL:X79787; NID:q1050468; PIDN:CA56185.1; PID:q1050471  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617  
A:Accession: C65047  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-136 <BLAT>  
A:Cross-references: GB:AE000352; GB:U00096; NID:q1789024; PIDN:AACT5721.1; PID:q1789029 A:Experimental source: strain K-12, substrain MG1655  
C:Superfamily: Bacillus subtilis conserved hypothetical protein yosm

Query Match 90.9%; Score 30; DB 1; Length 136;  
Best Local Similarity 83.3%; Pred. No. 1.4e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGGTSG 6  
|||||  
Db 51 GGGTAG 56

RESULT 25  
D85915  
hypothetical protein nrdi [imported] - Escherichia coli (strain O157:H7)  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001  
C:Accession: D85915  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew

iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoculis, K.; Apoda Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: D85915  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-136 <STO>  
A:Cross-references: GB:AE005174; NID:q12517113; PIDN:AA657784.1; GSPDB:GN00145; UMGCP A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: nrdi  
C:Superfamily: Bacillus subtilis conserved hypothetical protein yosm

Query Match 90.9%; Score 30; DB 2; Length 136;  
Best Local Similarity 83.3%; Pred. No. 1.4e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGGTSG 6  
|||||  
Db 51 GGGTAG 56

RESULT 26  
T46122  
30S ribosomal protein S9-like - Arabidopsis thaliana  
N:Alternate names: protein T2J13.80  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 18-Feb-2000  
C:Accession: T46122  
R:Rieger, M.; Gabel, C.; Mueller-huer, S.; Schefer, M.; Zipp, M.; Mewes, H.W.; Mayer submitted to the Protein Sequence Database, November 1999  
A:Reference number: Z23023  
A:Accession: T46122  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-143 <RIE>  
A:Cross-references: EMBL:AL132967  
A:Experimental source: cultivar Columbia; BAC clone T2J13  
C:Genetics:  
A:Map position: 3  
A:Introns: 105/1; 114/1  
A:Note: T2J13.80  
C:Superfamily: Escherichia coli ribosomal protein S9

Query Match 90.9%; Score 30; DB 2; Length 143;  
Best Local Similarity 83.3%; Pred. No. 1.5e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGGTSG 6  
|||||  
Db 82 GGGTTG 87

RESULT 27  
F84182  
hypothetical protein vng02219h [imported] - Halobacterium sp. NRC-1  
C:Species: Halobacterium sp. NRC-1  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: F84182  
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky R.; Leitauer, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja Jung, K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.; A:Title: Genome sequence of Halobacterium species NRC-1.  
A:Reference number: A84160; MUID:20504483  
A:Accession: F84182  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-185 <STO>

A:Cross-references: GB:AE004437; NID:q10579866; PIDN:AA618826.1; GSPDB:GN00138  
C:Genetics:  
A:Gene: VNC0219H

Query Match 90.9%; Score 30; DB 2; Length 185;  
Best Local Similarity 83.3%; Pred. No. 1.9e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGTSG 6  
||||:|  
DQ 22 GCGTAG 27

RESULT 28  
S56827  
conserved hypothetical protein YJL055w - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein J1148  
C:Species: Saccharomyces cerevisiae  
C:Date: 08-Jul-1995 #sequence\_revision 08-Sep-1995 #text\_change 08-Oct-1999  
C:Accession: S56827  
R:Pohl, T.M.; Allinovic, G.  
Submitted to the Protein Sequence Database, September 1995  
A:Reference number: S56793  
A:Accession: S56827  
A:Molecule type: DNA  
A:Residues: 1-245 <TOV>  
A:Cross-references: EMBL:Z49330; NID:q1008194; PIDN:CAA89346.1; PID:q1008195; MIPS:YJL055  
C:Genetics:  
A:Map position: 10L  
C:Superfamily: yeast conserved hypothetical protein YJL055w

Query Match 90.9%; Score 30; DB 2; Length 245;  
Best Local Similarity 83.3%; Pred. No. 2.5e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGTSG 6  
||||:|  
DQ 56 GCGTTG 61

RESULT 29  
T03371  
glycine-rich protein grp3 - maize  
C:Species: Zea mays (maize)  
C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 21-Jul-2000  
R:Goddemeier, M.L.; Wulff, D.; Felix, G.  
Plant Mol. Biol. 36, 799-802, 1998  
A:Title: Root-specific expression of a Zea mays gene encoding a novel glycine-rich protein  
A:Reference number: Z14904; MUID:98187261  
A:Accession: T03371  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-256 <GOD>  
A:Cross-references: EMBL:Y07781; NID:q1532070; PIDN:CAA69104.1; PID:q1532071  
A:Experimental source: strain DK105; root  
C:Genetics:  
A:Gene: grp3  
C:Superfamily: glycine-rich cell wall structural protein 1

Query Match 90.9%; Score 30; DB 2; Length 256;  
Best Local Similarity 83.3%; Pred. No. 2.6e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGTSG 6  
||||:|  
DQ 92 GCGTNG 97

RESULT 30

F75300  
hypothetical protein - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000  
C:Accession: F75300  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.  
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Ulterback, T.; Zalewski, C.;  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896  
A:Accession: F75300  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-276 <WHI>  
A:Cross-references: GB:AE002055; GB:AE000513; NID:q6460024; PIDN:AA611780.1; PID:q646  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DR2232  
A:Map position: 1

Query Match 90.9%; Score 30; DB 2; Length 276;  
Best Local Similarity 83.3%; Pred. No. 2.8e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGTSG 6  
||||:|  
DQ 61 GCGTAG 66

RESULT 31  
G85887  
hypothetical protein eutJ [imported] - Escherichia coli (strain O157:H7)  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001  
C:Accession: G85887  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May  
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Diallant, E.; Potamousis, K.; Apoda  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: G85887  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-278 <SPO>  
A:Cross-references: GB:AE005174; NID:q12516826; PIDN:AA657563.1; GSPDB:GN00145; UWGP:  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: eutJ

Query Match 90.9%; Score 30; DB 2; Length 278;  
Best Local Similarity 83.3%; Pred. No. 2.8e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGTSG 6  
||||:|  
DQ 145 GCGTTG 150

RESULT 32  
E65020  
ethanolamine utilization protein EutJ - Escherichia coli (strain K-12)  
C:Species: Escherichia coli  
C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 08-Oct-1999  
C:Accession: E65020  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;  
A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617  
A:Accession: E65020

A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-278 <BLAT>  
A:Cross-references: GB:AE000332; GB:U00096; NID:g1788789; PIDN:AAC75507.1; PID:g1788796;  
A:Experimental source: strain K-12, substrain MG1655  
C:Genetics:  
A:Gene: eutJ

Query Match  
Best Local Similarity 90.9%; Score 30; DB 2; Length 278;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
||||:|  
DB 145 GCGTTG 150

RESULT 33  
JCG502  
pectate lyase (EC 4.2.2.2) - Amycolata sp.  
C:Species: Amycolata sp.  
C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 21-Jul-2000  
C:Accession: JCG502  
R:Bruehlmann, F.; Keen, N.T.  
Gene 202, 45-51, 1997  
A:Title: Cloning, sequence and expression of the pel gene from an Amycolata sp.  
A:Reference number: JCG502; MUID:98087416  
A:Accession: JCG502  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-310 <BRD>  
A:Cross-references: GB:AF002241; NID:g2196706; PIDN:AAC38059.1; PID:g2196707  
C:Keywords: carbon-oxygen lyase

Query Match  
Best Local Similarity 90.9%; Score 30; DB 2; Length 310;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
||||:|  
DB 38 GCGTTC 43

RESULT 34  
T20807  
hypothetical protein FL3A7.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T20807  
R:McMurray, A.  
submitted to the EMBL Data Library, March 1997  
A:Reference number: Z19327  
A:Accession: T20807  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-340 <WIL>  
A:Cross-references: EMBL:Z93377; PIDN:CAB07573.1; GSPDB:GN00023; CESP:FL3A7.1  
A:Experimental source: clone FL3A7  
C:Genetics:  
A:Gene: CESP:FL3A7.1  
A:Map position: 5  
A:Introns: 6/2; 44/2; 80/3

Query Match  
Best Local Similarity 90.9%; Score 30; DB 2; Length 340;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
||||:|  
DB 224 GCGTAG 229

RESULT 35  
T34954  
probable UDP-N-acetylglucosamine-N-acetyluramyl-(pentapeptide) pyrophosphoryl-undeca

C:Species: Streptomyces coelicolor  
C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 21-Jan-2000  
C:Accession: T34954  
R:Saunders, D.C.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M  
submitted to the EMBL Data Library, August 1999  
A:Reference number: Z21563  
A:Accession: T34954  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-364 <SAU>  
A:Cross-references: EMBL:AL109663; PIDN:CAB51993.1; GSPDB:GN00070; SCOEDB:SC4A10.17C  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: murg; SCOEDB:SC4A10.17C  
C:Superfamily: murg protein

Query Match  
Best Local Similarity 90.9%; Score 30; DB 2; Length 364;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
||||:|  
DB 7 GCGTAG 12

RESULT 36  
I57555  
C-Maf protein - mouse  
C:Species: Mus sp. (mouse)  
C:Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 16-Jul-1999  
C:Accession: I57555  
R:Kurschner, C.; Morgan, J.I.  
Mol. Cell. Biol. 15, 246-254, 1995  
A:Title: The maf proto-oncogene stimulates transcription from multiple sites in a pro  
A:Reference number: I57555; MUID:95097997  
A:Accession: I57555  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-370 <RES>  
A:Cross-references: GB:S74567; NID:9807054; PIDN:AAB32820.1; PID:9807055  
C:Genetics:  
A:Gene: C-maf  
C:Superfamily: maf transforming protein; maf homology  
F:259-348/Domain: maf homology <MAF>

Query Match  
Best Local Similarity 90.9%; Score 30; DB 2; Length 370;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
||||:|  
DB 233 GCGTAG 238

RESULT 37  
B34261  
alanine dehydrogenase (EC 1.4.1.1) - Bacillus stearothermophilus  
C:Species: Bacillus stearothermophilus  
C:Date: 01-Jun-1990 #sequence\_revision 01-Jun-1990 #text\_change 11-Jun-1999  
C:Accession: B34261  
R:Kuroda, S.; Tanizawa, K.; Sakamoto, Y.; Tanaka, H.; Soda, K.  
Biochemistry 29, 1009-1015, 1990  
A:Title: Alanine dehydrogenases from two Bacillus species with distinct thermostabili  
ty and dehydrogenases  
A:Reference number: B34261; MUID:90254112  
A:Accession: B34261

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-372 <KUR>  
A:Cross-references: GB:M33299; NID:g142454; PIDN:AAA22211.1; PID:g142455  
C:Superfamily: alanine dehydrogenase; alanine dehydrogenase homology  
C:Keywords: NAD; oxalotransferase  
F:1-277/Domain: alanine dehydrogenase homology <ALA>  
F:170-198/Region: beta-alpha-beta NAD nucleotide-binding fold

Query Match 90.9%; Score 30; DB 2; Length 372;  
Best Local Similarity 83.3%; Pred. No. 3.7e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
|||||:1  
DB 175 GCGTAG 180

RESULT 38  
B46233  
transcription factor lmx-1 - golden hamster  
C:Species: Mesocricetus auratus (golden hamster)  
C>Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 16-Jul-1999  
C:Accession: B46233  
R:German, M.S.; Wang, J.; Chadwick, R.B.; Rutter, W.J.  
Gene Dev. 6, 2165-2176, 1992  
A:Title: Synergistic activation of the insulin gene by a LIM-homeo domain protein and a  
A:Reference number: A46233; MUID:93051335  
A:Accession: B46233  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-382 <GER>  
A:Cross-references: GB:X11406; NID:9587460; PIDN:CAA57163.1; PID:9587461  
A:Experimental source: Insulinoma cell line HIT-T-15 M2.2.2  
A>Note: sequence extracted from NCBI Backbone (NCBIN:117927, NCBI:117928)  
C:Genetics:  
A:Gene: lmx-1  
C:Superfamily: transcription factor 1sl-1; homeobox homology; LIM metal-binding repeat  
C:Keywords: DNA binding; duplication; homeobox; nucleos; transcription regulation; zinc  
F:94-147/Domain: LIM metal-binding repeat homology <LIM2>  
F:196-252/Domain: homeobox homology <HOX>

Query Match 90.9%; Score 30; DB 1; Length 382;  
Best Local Similarity 83.3%; Pred. No. 3.8e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
|||||:1  
DB 277 GCGTAG 282

RESULT 39  
T48109  
hypothetical protein F16M2.50 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
C:Accession: T48109  
R:Rieger, M.; Mueller-Auer, S.; Zipp, M.; Schaefer, M.; Mewes, H.W.; Rudd, S.; Lemcke, K.  
submitted to the Protein Sequence Database, April 2000  
A:Reference number: 224459  
A:Accession: T48109  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-382 <RIE>  
A:Cross-references: EMBL:AL138648  
A:Experimental source: cultivar Columbia; BAC clone F16M2  
C:Genetics:  
A:Map position: 3  
A:Introns: 313/3  
A>Note: F16M2.50

Query Match 90.9%; Score 30; DB 2; Length 382;  
Best Local Similarity 83.3%; Pred. No. 3.8e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
|||||:1  
DB 35 GCGTTG 40

RESULT 40  
D69809  
spore germination protein homolog yfkr - Bacillus subtilis  
C:Species: Bacillus subtilis  
C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 15-Oct-1999  
C:Accession: D69809  
R:Kunst, F.; Ogasawara, N.; Koszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber  
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;  
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fedret, C.; Ferrati,  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Gal  
lechi, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M  
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau  
Y., M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scani  
A:Authors: Schleich, S.; Schroeter, R.; Scofield, F.; Sekiguchi, J.; Sekowski, A.; Se  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida  
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis  
A:Reference number: A69580; MUID:98044033  
A:Accession: D69809  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-384 <KUN>  
A:Cross-references: GB:299108; GB:AL009126; NID:92633055; PIDN:CAB12607.1; PID:el1827  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: yfkr

Query Match 90.9%; Score 30; DB 2; Length 384;  
Best Local Similarity 83.3%; Pred. No. 3.8e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCGTSG 6  
|||||:1  
DB 64 GCGTAG 69

Search completed: February 4, 2002, 08:00:52  
Job time: 36 sec

---

**THIS PAGE BLANK (USPTO)**